Coupled Integral PINN for Discontinuity

Anonymous Authors[1]

Abstract

Physics-Informed Neural Networks (PINNs) solve forward PDEs by minimizing residual losses from the governing equations with initial and boundary conditions, but they often struggle with discontinuities such as shocks. In contrast, finite volume methods (FVM) handle discontinuities by enforcing integral conservation, which admits weak solutions. Motivated by this, we propose a *Coupled Integral PINN (CI-PINN)* that augments a standard PINN with an auxiliary network for integral potentials and coupled integral constraints. This improves robustness near shocks while avoiding meshing and the numerical flux integration/reconstruction used in classical schemes. We validate CI-PINN on forward benchmarks including Burgers, Buckley–Leverett, the Euler system, and the Shallow-Water equations.

1. Introduction

Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019; 2017) have become a central building block in scientific machine learning (SciML), providing a flexible way to fuse neural function approximation with the structure of governing differential equations (Karniadakis et al., 2021; Lu et al., 2021). PINN embeds the residual of a partial differential equation (PDE) into the training objective via automatic differentiation. While PINNs have proven effective in modeling diverse physical systems, they exhibit well-documented limitations when the underlying solution exhibits sharp fronts or discontinuities (Krishnapriyan et al., 2021; Fuks & Tchelepi, 2020). (Despite PINN's success in many application areas, applying it to problems involving sudden changes or discontinuities presents significant challenges (Karniadakis et al., 2021).) In this work we try to theoretically investigate why PINN fails on PDE systems with discontinuities. To better understand the challenge these PDEs impose on PINN, we leave out the data loss in PINN and therefore focus on the forward problem without interior observational data. Our goal is to investigate why PINN cannot approximate PDE solutions directly from the governing equations and prescribed initial/boundary conditions for systems with discontinuities, and propose an approach that addresses the issue.

We argue that the root cause is a fundamental mismatch between the *strong-form residual objective* and the *hyperbolic nature of the target solution*, using a viewpoint from an optimization and function-approximation perspective. PINN training inherently fails to converge to the true solution in the presence of discontinuities. We formalize this failure mode through two key theoretical insights that PINN strong-form is not identifiable, and that PINN physics loss diverges near shocks.

Classical numerical analysis does not have this issue. This is because these regimes are handled by enforcing PDE structure through integral flux-balance principles (e.g., finite volume methods), which remain meaningful at shocks. This motivates us to propose integral-form PINN constraints that extend PINNs to shock-dominated problems, aiding strong-form PINN objectives so that PINN can behave in the discontinuous regime.

1.1. Our Contribution

In this work, we theoretically investigate why vanilla PINNs fail in discontinuity regime. Given the lessons, we propose to enforce the conservation law in its integral form via auxiliary networks, we naturally handle discontinuities without derivative singularities. We name our architecture **Coupled Integral PINN (CI-PINN)**, which explicitly decouples the state variable from its integral representation. Our specific contributions are:
(1) **Optimization analysis:** We provide a theoretical analysis of why strong-form PINNs fail at shocks, proving the existence of an optimization barrier that penalizes convergence to the entropy solution.
(2) **Algorithm Design:** We introduce a dual-network architecture that enforces the integral conservation law directly, ensuring that the loss function remains bounded and smooth even in the presence of sharp shocks, meanwhile because CI-PINN avoids any explicit local or global numerical quadrature, it can be seamlessly integrated with existing PINN

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Submitted to the AI for Science workshop (ICML 2026).

enhancements—such as domain decomposition, adaptive weighting, or curriculum sampling—without additional algorithmic overhead.
(3) **Empirical Performance:** We demonstrate that CI-PINN significantly outperforms standard PINN variants on benchmark hyperbolic systems (Burgers, Buckley-Leverett, Euler), effectively capturing sharp shock interfaces and accurately recovering the ground truth entropy solution where traditional PINNs fail.

The proposed CI-PINN offers three principal advantages:

(1) Better accuracy. On benchmark hyperbolic conservation laws, CI-PINN captures both shock and rarefaction waves accurately, and achieve best accuracy on most of evaluation benchmark.
(2) Automatic conservation. By enforcing a pointwise integral form of the PDE during training, the network simultaneously approximates the integral, local, and global conservation statements once convergence is reached.
(3) Algorithmic flexibility. CI-PINN avoids any explicit local or global numerical quadrature. The loss are evaluated on mesh-less point, it can be seamlessly integrated with existing PINN enhancements without additional algorithmic overhead.

The remainder of this paper is organized as follows. Section 2 introduces the mechanism of the shock waves, then delivers an argument why vanilla PINNs almost fail at every discontinuity and how CI-PINN overcome this. Section 3 present the proposed dual network architecture and how the consistency between the neural networks and the true physical solution are enforced. Section 4 presents experimental results on classical benchmark problems, highlighting the advantages of Coupled Integral PINNs over other methods. Finally, Section **??** concludes and outlines future work.

1.2. Related Work

The difficulty of applying strong-form PINNs to discontinuous solutions has been widely recognized. For inviscid Burgers' equation, standard PINNs often converge to overly smooth surrogates rather than the correct shock profile, and similar failures occur in more complex hyperbolic settings with moving shocks and compound interactions (Diab & Kobaisi, 2021). These limitations affect *both* forward and inverse problems: while additional data may partially constrain inverse solutions, it does not resolve the fundamental mismatch between pointwise residual minimization and the admissible weak/entropy solution concept (Krishnapriyan et al., 2021; Fuks & Tchelepi, 2020).

A large body of work attempts to improve training stability without changing the strong-form objective. Domain decomposition methods (e.g., XPINNs and related multi-network or overlapping-subdomain strategies) localize approximation and improve scalability (Jagtap & Karniadakis, 2020; Jagtap et al., 2020). Adaptive or clustered residual sampling increases collocation density in hard regions (often near shocks) (Mao et al., 2020), and gradient-based weighting further rebalances training signals around discontinuities (Liu et al., 2024). Curriculum/causal schedules and architectural modifications address temporal stiffness and gradient pathologies (Wang et al., 2022a; 2020; 2022b). Although these approaches can reduce variance and improve robustness, they typically do *not* remove the core issue that the strong-form residual becomes a poor learning signal as interfaces sharpen, leading to smeared shocks or distorted plateau levels.

Another direction replaces pointwise residuals with weak/variational formulations that are better aligned with discontinuous solution concepts. Examples include Ritz–Galerkin and related variational solvers (Yu et al., 2018), mortar/domain-interface formulations (Jagtap & Karniadakis, 2020), and Petrov–Galerkin or hp-variational PINNs (Kharazmi et al., 2021b;a). Weak-PINNs further target weak residuals and error control (Ryck et al., 2022; Chaumet & Giesselmann, 2023). These methods soften derivative singularities by shifting differentiation onto smoother test functions, but their effectiveness depends on test-space design and numerical quadrature accuracy; moreover, for nonlinear conservation laws, weak satisfaction alone does not guarantee uniqueness without an admissibility mechanism (e.g., entropy selection), and multidimensional implementations can be computationally heavy.

In the specific context of conservation laws, many methods enforce integral flux balance by discretizing space–time and applying numerical quadrature (e.g., trapezoidal rules or finite-volume-style control volumes) during training (Mei et al., 2024; Jeon & Kim, 2021; Patel et al., 2022; Li et al., 2023; 2024a). These approaches substantially improve shock consistency, and cvPINN (Patel et al., 2022) has become a representative *standard* baseline for conservation enforcement via local control volumes and numerical quadrature. However, such formulations can require careful discretization, quadrature, and sometimes reconstruction choices, and they primarily enforce local balance rather than directly learning a global integral representation.

Our CI-PINN targets the *forward* solution of nonlinear conservation laws in a mesh-free manner by fitting an auxiliary potential network and enforcing coupled integral constraints without explicit space–time discretization, quadrature, or flux reconstruction. This also distinguishes our setting from supervised operator learning (e.g., DeepONet (Lu et al., 2019) and neural operators (Li et al., 2020; Gupta et al., 2021)) and hybrid operator learners (e.g., PINO (Li et al., 2024b; Goswami et al., 2023)) and their variants, since we focus on unsupervised physics-based forward solving

without interior observational data. In experiments, we compare against general optimization enhancements (Adaptive Learning Rate (Wang et al., 2023), Causal Training (Wang et al., 2022a), Modified MLP (Wang et al., 2020), Random Weight Factorization (Wang et al., 2022b)) and conservation-specific formulations including cvPINN (Patel et al., 2022), IPINN (Rajvanshi & Ketcheson), and a small-scale comparison with Bayesian ProbConserv (Hansen et al., 2024) in H.

2. Failure of PINNs for Discontinuous PDE Solutions in $L^2(\Omega)$

We analyze a fundamental failure mode of strong-form PINNs when the target solution is discontinuous. In this section we will use $\mathbf{q}$ instead of the typical $\mathbf{u}$ to represent conservative quantities to split out from the primitive element in section 3

2.1. Preliminaries: Why shocks form

Shocks arise most naturally in *nonlinear hyperbolic conservation laws*, i.e., PDEs that admit the conservative form $\mathbf{q}_t + \nabla\cdot \mathbf{F}(\mathbf{q}) = 0$ (or $q_t + f(q)_x = 0$) and exhibit wave-like propagation. For example, in the scalar conservation law $q_t + f(q)_x = 0$, smooth solutions propagate along characteristic curves with speed $f'(q)$. When $f'(q)$ varies with the state, different parts of the wave travel at different speeds, leading to compression: faster characteristics overtake slower ones, the profile steepens, and the gradient q_x blows up in finite time. Physically, this corresponds to wave steepening in compressible media, where high-pressure (or high-density) regions propagate faster and "pile up" into a thin dissipative layer, producing an irreversible jump. Beyond this point, classical (strong) solutions cease to exist and the physically relevant solution must be interpreted in the weak sense(integral). Conservation laws appear broadly in science and engineering, including fluid dynamics, transport, and porous-media flow. Their integral formulation is obtained by integrating the differential form over a control volume $\Omega \subset \mathbb{R}^d$ and applying the divergence theorem:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega} q(\mathbf{x},t)\,\mathrm{d}\Omega \;=\; -\int_{\partial\Omega} \mathbf{F}(q)\cdot\mathbf{n}\,\mathrm{d}\Gamma\,, \tag{1}$$

where $\partial\Omega$ is the boundary of Ω and $\mathbf{n}$ denotes the outward unit normal. A detailed derivation is provided in Appendix C.

2.2. PDE and strong-form PINN objective

Let $\mathbf{q} : \Omega \times (0,T) \to \mathbb{R}^m$ be the unknown field on $\Omega \subset \mathbb{R}^d$. We consider nonlinear hyperbolic PDEs that admit discontinuous physical solutions, written in a standard flux-form operator where $\mathbf{F}$ is the flux operator:

$$\partial_t \mathbf{q} + \nabla\cdot\mathbf{F}(\mathbf{q}) = \mathbf{0}, \qquad \mathbf{F}(\mathbf{q}) \in \mathbb{R}^{m\times d}. \tag{2}$$

Strong-form PINNs typically minimize the squared residual

$$\begin{aligned}\mathcal{L}_{\text{strong}}(\mathbf{q}_\theta) = \|\partial_t \mathbf{q}_\theta + \nabla\cdot\mathbf{F}(\mathbf{q}_\theta)\|^2_{L^2(\Omega_T)}\,,\\ \Omega_T := \Omega\times(0,T),\end{aligned} \tag{3}$$

where $\mathbf{q}_\theta$ is a smooth neural approximation.

2.3. Core pathologies of strong-form training

Non-Uniqueness of Strong-Form Minimizers: Since the true shock contains a derivative singularity, the strong-form loss is unbounded at the physical solution and have very large loss on the approximating solution. Consequently, neural networks are biased toward spurious, smooth approximations that yield deceptively low residuals but incorrect physics (see Proposition 2.1). From optimization perspective, this means physical solution is not global minimum.

The Optimization Barrier: We reveal that the standard PINN loss is "inverse-consistent" near a shock—as a candidate solution approaches the true sharp interface, the residual loss strictly *diverges*. This implies that the true solution is not a local minimum in the standard loss landscape; rather, the gradient actively repels the optimizer away from the shock (see Proposition 2.3).

Proposition 2.1 (Spurious smooth low-loss solutions dominate the strong-form objective)**.** *Let $\mathbf{q}^*$ be a physically relevant discontinuous weak solution of* (2) *containing at least one jump discontinuity. Consider minimizing the strong-form residual objective over a smooth hypothesis class:*

$$\mathcal{L}_{\text{strong}}(\tilde{\mathbf{q}}) := \|\partial_t \tilde{\mathbf{q}} + \nabla\cdot\mathbf{F}(\tilde{\mathbf{q}})\|^2_{L^2(\Omega_T)}\,, \qquad \tilde{\mathbf{q}} \in C^1(\Omega_T).$$

Then there exist smooth functions $\tilde{\mathbf{q}}$ with arbitrarily small *strong-form loss even though $\tilde{\mathbf{q}}$ does not reproduce the discontinuous structure of $\mathbf{q}^*$ (e.g., it yields an overly smeared shock/front). Consequently, minimizing $\mathcal{L}_{\text{strong}}$ is non-selective within smooth classes and can favor diffused, non-physical smooth states.*

Remark 2.2. Proposition 2.1 is an *existence / non-selectivity* statement: a small (even vanishing) interior strong-form residual does not certify that the learned solution matches the physically relevant discontinuous weak solution. Intuitively, the residual contains spatial derivatives; if a shock/front is spread over a wider layer, the gradients shrink, and an L^2 objective can be made small even though the interface is severely smeared.

Proposition 2.3 (Strong-form blow-up as a discontinuity sharpens)**.** *Let $\mathbf{q}^*$ be a discontinuous physical solution containing a jump of amplitude $\Delta\mathbf{q} \neq \mathbf{0}$ across an interface. Let $\{\mathbf{q}^\varepsilon\}_{\varepsilon>0}$ be a family of smooth approximations that*

resolves the jump over a transition layer of thickness ε. Assume $\mathbf{F}$ is continuously differentiable and its Jacobians remain bounded on the relevant state range. Then the strong-form residual scales as

$$\mathcal{L}_{\text{strong}}(\mathbf{q}^\varepsilon) = \|\partial_t \mathbf{q}^\varepsilon + \nabla \cdot \mathbf{F}(\mathbf{q}^\varepsilon)\|^2_{L^2(\Omega_T)} \gtrsim \frac{C}{\varepsilon},\ \varepsilon \to 0, \tag{4}$$

*for some $C > 0$ depending on $\Delta\mathbf{q}$ and $\mathbf{F}$. Consequently, sharpening an approximation toward the correct discontinuity (*smaller ε*)* increases *the strong-form penalty.*

Proof. When a fixed jump of magnitude $\|\Delta\mathbf{q}\|$ between two levels must occur within a transition layer of thickness ε, the gradient (slope) of this transition necessarily scales as $\|\Delta\mathbf{q}\|/\varepsilon$. That is, within the $\mathcal{O}(\varepsilon)$-thick transition region, gradients scale as $\|\nabla\mathbf{q}^\varepsilon\| \sim \mathcal{O}\Big(\frac{\|\Delta\mathbf{q}\|}{\varepsilon}\Big)$. By the chain rule,

$$\nabla \cdot \mathbf{F}(\mathbf{q}^\varepsilon) = \sum_{j=1}^{d} \partial_{x_j} \mathbf{F}_j(\mathbf{q}^\varepsilon) = \sum_{j=1}^{d} \mathbf{J}_{\mathbf{F}_j}(\mathbf{q}^\varepsilon)\, \partial_{x_j} \mathbf{q}^\varepsilon,$$

so bounded Jacobians imply $\|\nabla \cdot \mathbf{F}(\mathbf{q}^\varepsilon)\| \sim \mathcal{O}(1/\varepsilon)$ inside the layer. Hence, $\|\partial_t \mathbf{q}^\varepsilon + \nabla \cdot \mathbf{F}(\mathbf{q}^\varepsilon)\| \sim \mathcal{O}(1/\varepsilon)$ over a region of spatial measure $\mathcal{O}(\varepsilon)$ and time measure $\mathcal{O}(1)$, yielding $\mathcal{L}_{\text{strong}}(\mathbf{q}^\varepsilon) \gtrsim (\varepsilon) \cdot \big(\frac{1}{\varepsilon}\big)^2 = \frac{1}{\varepsilon}$. □

This completes the proof of proposition 2.3 and 2.1, we provide a detailed proof sketch in appendix A

Corollary 2.4 (The Optimization barrier)**.** *Combining Propositions 2.1 and 2.3 establishes a fundamental pathology in the strong-form optimization landscape. Specifically, $\mathcal{L}_{\text{strong}}$ is* inverse-consistent *with discontinuity sharpness:* ***Divergence at the Solution** ($\varepsilon \to 0$)**:** As the approximation $\mathbf{q}^\varepsilon$ converges to the physical sharp shock $\mathbf{q}^*$, the loss diverges: $\mathcal{L}_{\text{strong}} \to \infty$.* ***Convergence to Artifacts** ($\varepsilon \to \infty$)**:** As the approximation deviates from physics via excessive smoothing, the loss converges to a global minimum: $\mathcal{L}_{\text{strong}} \to 0$.*

This creates an optimization barrier: minimizing the strong-form residual inherently penalizes physical accuracy and drives the optimizer toward non-physical, smeared solutions.

Remark 2.5. (Adding an initial-condition penalty does not remove the strong-form pathology). Adding an initial-condition (IC) loss anchors the state at $t = 0$, but it does not alter the interior scaling mechanism in Proposition 2.3: for any smooth approximation that must realize a fixed jump by a transition layer of thickness ε, the strong-form contribution still satisfies $\mathcal{L}_{\text{strong}}(\mathbf{q}^\varepsilon) \gtrsim \frac{1}{\varepsilon} \quad (\varepsilon \to 0)$. Hence IC supervision does not remove the blow-up barrier associated with sharpening discontinuities, and the optimizer may still prefer overly diffused profiles at later times.

2.4. Why the Coupled Integral PINN resolves the failure mode

Remark 2.6 (How CI-PINN avoids strong-form shock blow-up). The key difficulty for classical PINNs on hyperbolic problems is that a sharp jump in the state $\mathbf{q}$ forces derivatives to scale like $1/\varepsilon$ inside an $\mathcal{O}(\varepsilon)$ shock layer, which can make strong-form residual penalties grow as the discontinuity sharpens (Proposition 2.3). CI-PINN avoids this failure mode by enforcing the PDE through *coupled integral (potential) variables* rather than directly squaring shock-amplified derivatives of the primitive/state fields.

Concretely, we learn a primitive network $\mathbf{u}_\theta$ and a potential network $\mathbf{S}_\phi = \{\mathbf{S}_{\phi,j}\}_{j=1}^d$, define $\mathbf{q}_\theta = \mathbf{q}(\mathbf{u}_\theta)$, and impose the coupled constraints

$$\mathbf{q}_\theta \approx \nabla \cdot \mathbf{S}_\phi, \qquad \partial_t \mathbf{S}_{\phi,j} + \mathbf{F}_j(\mathbf{q}_\theta) \approx \mathbf{0}, \quad j = 1, \ldots, d,$$

which correspond exactly to the *coupled-loss* and *physical-loss* terms in our objective. Because $\mathbf{S}_\phi$ acts as an antiderivative of $\mathbf{q}_\theta$ (through $\nabla \cdot \mathbf{S}_\phi \approx \mathbf{q}_\theta$), a jump in $\mathbf{q}$ typically manifests as a much milder non-smoothness in $\mathbf{S}$ (often a continuous function with a kink), so the training signal does not require differentiating the discontinuous state itself. In turn, the dominant contributions in the physical-loss arise from $\partial_t \mathbf{S}_{\phi,j}$ and the flux evaluation $\mathbf{F}_j(\mathbf{q}_\theta)$, rather than from squaring $1/\varepsilon$-scale spatial derivatives of $\mathbf{q}$. This removes the main driver behind strong-form blow-up and makes shock sharpening compatible with optimization. Finally, other loss terms involved are enforced to be stable.

On the remaining kink in S. Even if $\mathbf{S}$ is only continuous with a corner at the shock, this non-smoothness is mild: $\mathbf{S}$ remains Lipschitz and is smooth almost everywhere, with the kink confined to a codimension-one interface (measure zero in Ω_T). As a result, L^2 losses involving only first derivatives of $\mathbf{S}$ stay finite and well-behaved, unlike strong-form losses on $\mathbf{q}$ whose gradients can scale like $\mathcal{O}(1/\varepsilon)$ when a jump is represented by a vanishing-thickness layer.

3. Methodology

3.1. The CI-PINN method

Inspired by the integral form and the corresponding FVM in numerical methods, the Coupled Integral PINN (CI-PINN) employs two fully connected neural networks (see Figure 1). The first network, $\tilde{\mathbf{u}}(x,t)$, approximates the *primitive* state variables of the solution (e.g., density, velocity, pressure in compressible flow). Since the governing PDE is typically written in conservative form, we define the corresponding *conserved* variables through a deterministic mapping

$$\mathbf{q}(x,t) = \mathbf{q}(\mathbf{u}(x,t)), \qquad \tilde{\mathbf{q}}(x,t) := \mathbf{q}(\tilde{\mathbf{u}}(x,t)).$$

The second network learns a potential (integral) representation $\tilde{\mathbf{S}}(x,t) = \{\tilde{\mathbf{S}}_j(x,t)\}_{j=1}^d$ associated with the conserved

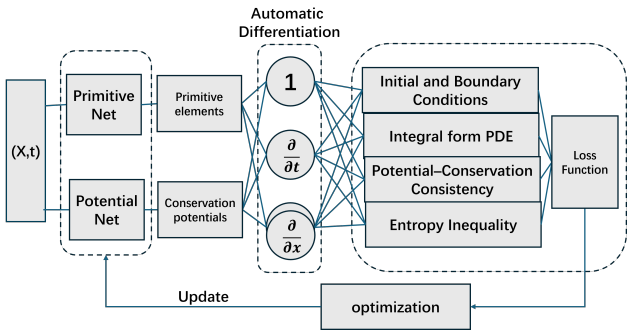


Figure 1. CI-PINN architecture to solve the forward problem of nonlinear PDEs.

quantities. In the general d-dimensional setting, CI-PINN enforces the coupled relations

$$\tilde{\mathbf{q}}(x,t) \approx \nabla \cdot \tilde{\mathbf{S}}(x,t),$$

$$\partial_t \tilde{\mathbf{S}}_j(x,t) + \mathbf{F}_j(\tilde{\mathbf{q}}(x,t)) \approx \mathbf{0}, \quad j = 1, \ldots, d,$$

which correspond to an integral-form relaxation of the conservation law $\partial_t \mathbf{q} + \nabla \cdot \mathbf{F}(\mathbf{q}) = \mathbf{0}$. The required derivatives (e.g., $\nabla \cdot \tilde{\mathbf{S}}$ and $\partial_t \tilde{\mathbf{S}}$) are computed through automatic differentiation. Because differentiation is applied to the learned potential rather than directly to discontinuous primitive states, the proposed method remains compatible with shock solutions while retaining a simple neural architecture.

As a special case in 1D scalar conservation laws, the primitive and conserved variables coincide ($\mathbf{q} = u$). The potential reduces to a scalar function $S(x,t)$ defined as the spatial antiderivative of the solution, implying $u(x,t) = \partial_x S(x,t)$. For instance, in the inviscid Burgers' equation (where flux $f(u) = u^2/2$), the **Primitive Net** maps spacetime coordinates (x,t) to the state variable u, while the **Potential Net** maps (x,t) to the potential S (the integral of the conserved quantity). These networks are trained jointly to enforce consistency ($\partial_x S \approx u$) and the integral conservation law ($\partial_t S + u^2/2 \approx 0$).

Training CI-PINN amounts to enforcing four conditions that mirror the four constraint blocks in Figure 1. First, the primitive network output $\tilde{\mathbf{u}}(x,t)$ must satisfy the prescribed initial and boundary data. Second, the learned potential $\tilde{\mathbf{S}}(x,t)$ is required to be consistent with the conserved variables through $\nabla \cdot \tilde{\mathbf{S}}(x,t) \approx \tilde{\mathbf{q}}(x,t)$, where $\tilde{\mathbf{q}} = \mathbf{q}(\tilde{\mathbf{u}})$. Third, conservation is enforced in integral form by requiring $\partial_t \tilde{\mathbf{S}}_j(x,t) + \mathbf{F}_j(\tilde{\mathbf{q}}(x,t)) \approx \mathbf{0}$ for $j = 1, \ldots, d$. Finally, to select the physically admissible weak solution in the presence of shocks, we impose the entropy inequality. When jointly enforced, these four constraints are sufficient to recover the unique entropy solution almost everywhere; removing the consistency constraint leads to architectural degeneracy while removing any of the other three leads to non-uniqueness.

Figure 1 illustrates the architecture of CI-PINN, where normalized spatial and temporal coordinates serve as inputs to the two MLPs. The outputs, $\tilde{\mathbf{u}}(x,t)$ and $\tilde{\mathbf{S}}(x,t)$, are optimized to satisfy 4 conditions mentioned above. Under the PINN setting, these four conditions in Figure 1 are realized by optimizing the summation of four losses.

The total loss L is given by:

$$\begin{aligned} L = \lambda_{\text{ibc}} & \underbrace{\frac{1}{N_b} \sum_{j=1}^{N_b} \left\| \mathbf{u}_{\text{IBC}}(x_j,t_j) - \tilde{\mathbf{u}}(x_j,t_j) \right\|_2^2}_{\text{initial-boundary loss}} \\ & + \lambda_{\text{phy}} \underbrace{\frac{1}{N_f\, d} \sum_{i=1}^{N_f} \sum_{k=1}^{d} \left\| \partial_t \tilde{\mathbf{S}}_k(x_i,t_i) + \mathbf{F}_k(\tilde{\mathbf{q}}(x_i,t_i)) \right\|_2^2}_{\text{integral-form (physical) loss}} \\ & + \lambda_{\text{cpl}} \underbrace{\frac{1}{N_f} \sum_{i=1}^{N_f} \left\| \nabla \cdot \tilde{\mathbf{S}}(x_i,t_i) - \tilde{\mathbf{q}}(x_i,t_i) \right\|_2^2}_{\text{coupling loss}} \\ & + \lambda_{\text{ent}} \underbrace{\frac{1}{N_f} \sum_{i=1}^{N_f} \Big[\max\big(0,\ \partial_t \eta(\tilde{\mathbf{q}}) + \nabla \cdot \boldsymbol{\phi}(\tilde{\mathbf{q}})\big) \Big]^2}_{\text{entropy admissibility loss}} \\ & + \lambda_{\text{str}} \underbrace{\frac{1}{N_f} \sum_{i=1}^{N_f} (1 - w_i) \left\| \partial_t \tilde{\mathbf{q}}(x_i,t_i) + \nabla \cdot \mathbf{F}(\tilde{\mathbf{q}}(x_i,t_i)) \right\|_2^2}_{\text{adaptive strong-form loss}}. \end{aligned} \tag{5}$$

Initial-Boundary Loss: This component ensures that the solution $\tilde{\mathbf{u}}(x,t)$ satisfies the initial and boundary conditions of the PDE,it minimizing the difference between $\tilde{\mathbf{u}}(x,t)$ and the true values at boundary and initial conditions; This loss term often referred to as the "data loss" in PINNs, aligns the predicted solution with the provided boundary and initial condition.

Physical Loss: This component enforces the integral-form PDE constraint through the potential variable by penalizing the mismatch in $\partial_t \tilde{\mathbf{S}}_j(x,t) + \mathbf{F}_j(\tilde{\mathbf{q}}(x,t))$. This formulation enforces conservation in integral form while avoiding direct differentiation of discontinuous primitive states. This necessitates the addition of the entropy loss to ensure the uniqueness and physical realism of the solution.

Coupled Loss: The coupled loss enforces the correct relationship between the conserved-state representation $\tilde{\mathbf{q}}(x,t) = \mathbf{q}(\tilde{\mathbf{u}}(x,t))$ and the potential $\tilde{\mathbf{S}}(x,t)$ by constraining $\nabla \cdot \tilde{\mathbf{S}}(x,t)$ to match $\tilde{\mathbf{q}}(x,t)$. It is not necessary to fix the absolute values of $\tilde{\mathbf{S}}(x,t)$; instead, the focus is on constraining its divergence, ensuring consistency between the conserved quantities and their potential representation.

Entropy Admissibility Loss: While the integral form of a hyperbolic conservation law enforces Rankine–Hugoniot jump conditions, it does not guarantee uniqueness and may admit non-physical weak solutions (Dafermos & Dafermos, 2005). To select the physically relevant solution, one imposes an entropy condition, e.g., via vanishing viscosity limits (LeVeque, 1992a; Bianchini & Bressan, 2005). In CI-PINN, we enforce entropy admissibility using a convex entropy function $\eta : \mathbb{R}^N \to \mathbb{R}$ and an associated entropy flux $\boldsymbol{\phi} : \mathbb{R}^N \to \mathbb{R}^d$, defined in the conserved variables $\mathbf{q}$. For each spatial direction $i \in \{1, \ldots, d\}$ with flux $\mathbf{F}^{(i)}(\mathbf{q})$, the entropy pair satisfies the compatibility relation $\nabla_{\mathbf{q}}\boldsymbol{\phi}^{(i)}(\mathbf{q}) = \nabla_{\mathbf{q}}\eta(\mathbf{q})^\top \nabla_{\mathbf{q}}\mathbf{F}^{(i)}(\mathbf{q})$. The entropy solution satisfies the entropy inequality $\partial_t \eta(\mathbf{q}) + \nabla \cdot \boldsymbol{\phi}(\mathbf{q}) \leq 0$.

Adaptive Strong Loss: While the strong-form residual is ill-posed at discontinuities (Section 2), it remains a highly informative and computationally efficient training signal in regions where the solution is smooth and classical derivatives exist. We therefore couple the differential and integral objectives through an adaptive shock mask $w_i \in [0, 1]$ and apply the strong-form penalty only where the local flow is estimated to be smooth.

Concretely, we construct w_i from a compression indicator based on the *negative spatial divergence* of the predicted velocity field: $w_i = \sigma\Big(k\,\mathrm{ReLU}\big(-\nabla_{\mathbf{x}} \cdot \tilde{\mathbf{v}}(\mathbf{x}_i, t_i)\big)\Big)$,$\sigma(z) = \frac{1}{1+e^{-z}}$. The term $\mathrm{ReLU}(-\nabla_{\mathbf{x}} \cdot \tilde{\mathbf{v}})$ activates only in compressive regions, which are well known to precede shock formation in hyperbolic dynamics. The sharpness parameter $k > 0$ controls the transition of the mask;

4. Experiments

Benchmarks & Metrics. We evaluate CI-PINN on canonical hyperbolic PDE benchmarks spanning both scalar laws and systems. Our 1D benchmarks include the inviscid Burgers' equation, the Buckley–Leverett (BL) equation, and the Euler shock-tube problems (Sod and Lax). We further consider two 2D benchmarks with complex multi-wave interactions and geometric structure: a 2D Euler Riemann problem and the 2D shallow-water equations (SWE).

To assess performance near discontinuities in *1D*, we partition the space–time domain $\Omega \times [0, T]$ into a *shock neighborhood* (Ω_s), a *rarefaction region* (Ω_r), and the *full domain* (Ω) (see Appendix G for definitions).

4.1. Main result on 1D Benchmarks

The quantitative results for all 1D benchmarks (Tables 1–4) suggest that in the majority of tested cases, CI-PINN achieves the lowest errors in shock and global regions while maintaining high accuracy in smooth rarefaction zones.

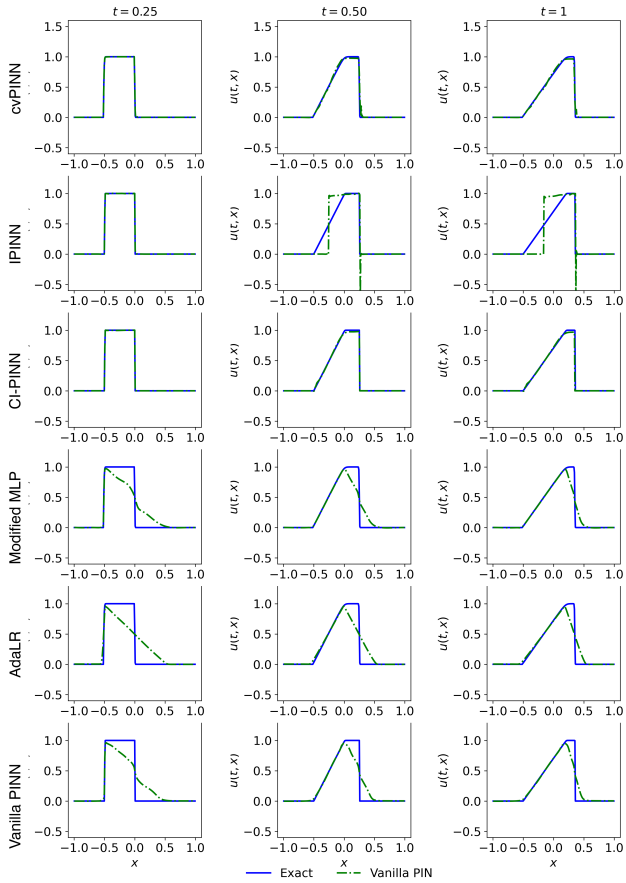


Figure 2. We utilize a square-wave initial condition ($u(x, 0) = 1$ for $x \in (-0.5, 0)$ as in the left-most panels), which naturally induces both a shock (where characteristics pile up, ie., sudden changes in right-most panels) and a rarefaction wave (where characteristics fan apart, ie., linear slope in right-most panels.)

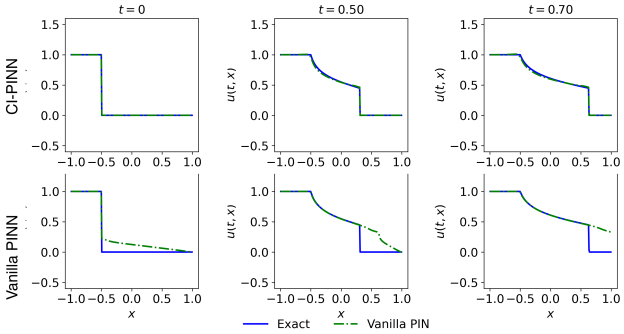
Shock and Rarefaction Analysis. Figure 2 highlights a known failure mode of strong-form training: the Vanilla PINN smooths the discontinuity and fails to represent the shock sharply. In contrast, integral-form training (CI-PINN, cvPINN, IPINN) is substantially more shock-consistent and yields sharper interfaces, consistent with enforcing conservation in a way that aligns with Rankine–Hugoniot-type jump behavior. However, integral formulations can be less stable in smooth rarefaction regions, where non-physical oscillations or biased profiles may appear if admissibility is not sufficiently controlled. Between the integral baselines, CI-PINN provides the best balance across regimes: it matches the shock location while reducing rarefaction-region artifacts, leading to the lowest global L_1 error in Table 1. Similar trends hold for the BL equation and the 1D Euler shock tubes (Tables 2–4).

Buckley–Leverett. In the Buckley–Leverett problem (Table 2), the non-convex flux $f(u)$ produces a compound wave in which a rarefaction and a shock appear in close proximity.

Table 1. Error metrics for Burgers equation across models (values in 10^{-3}).

Region	Metric	CI-PINN	cvPINN	IPINN	Mod. MLP	Causal	RWF	AdaLR	Vanilla
Global	L_2	**47.04**	96.15	153.23	120.62	282.44	131.00	146.20	123.87
	L_1	**7.55**	18.18	65.39	54.59	109.26	58.74	71.94	56.58
	L_∞	1012.71	997.44	1594.20	750.43	1268.73	875.97	**497.03**	915.87
Shock	L_2	**103.67**	211.89	166.24	258.13	561.66	280.65	307.54	266.31
	L_1	**21.26**	62.34	39.80	217.81	325.96	232.72	265.51	222.87
	L_∞	1012.71	997.44	1594.20	750.43	1262.88	875.97	**497.03**	915.87
Raref.	L_2	25.47	60.45	225.02	**7.40**	112.72	37.09	22.34	18.92
	L_1	10.95	19.54	159.59	**3.62**	99.60	5.91	14.89	7.45
	L_∞	1012.71	997.44	1000.35	390.70	**215.62**	875.97	374.24	915.87

Table 2. Error metrics for BL equation across models (values in 10^{-3}).



Region	Metric	CI-PINN	cvPINN	IPINN	Mod. MLP	Causal	RWF	AdaLR	Vanilla
Global	L_2	**17.15**	83.02	239.59	149.87	69.51	138.67	118.75	147.83
	L_1	**6.76**	26.76	133.14	80.28	35.96	64.16	60.65	78.17
	L_∞	444.08	794.39	1479.38	445.39	697.07	447.82	**440.21**	451.45
Shock	L_2	**28.64**	181.74	337.91	263.95	131.11	260.39	235.03	261.74
	L_1	**8.52**	91.77	227.89	182.55	94.32	180.45	163.71	180.62
	L_∞	444.08	794.39	1479.38	445.39	697.07	447.82	**440.21**	451.45
Raref.	L_2	25.07	42.75	420.00	**4.32**	70.88	7.13	16.95	5.68
	L_1	22.52	21.56	396.93	**2.24**	61.84	3.85	10.96	4.41
	L_∞	**57.16**	794.39	1479.38	71.89	304.96	171.17	128.16	62.54

Table 3. Averaged error metrics (across u, p, ρ) for Sod Shock Tube (values in 10^{-3}). Best is bold; second best is underlined. For full diagram please refer to appendix E

Region	Metric	CI-PINN	cvPINN	IPINN	Mod.MLP	Causal	RWF	AdaLR	Vanilla
Global	L_1	**6.77**	11.84	360.28	143.78	203.26	140.32	180.60	209.07
	L_2	**25.94**	32.04	558.38	201.35	369.30	200.31	255.16	392.55
	L_∞	946.43	1131.47	4378.85	**927.25**	1210.98	1094.87	1005.75	952.66
Shock	L_1	**15.36**	30.79	534.18	277.17	387.46	281.46	267.74	327.86
	L_2	**49.84**	61.07	698.27	324.45	523.87	330.20	321.00	465.65
	L_∞	946.43	1131.47	3994.40	**927.25**	1199.11	1094.87	1005.75	952.61
Raref.	L_1	**42.93**	46.35	660.08	171.76	217.06	184.58	297.09	480.75
	L_2	**76.78**	80.73	714.11	213.95	264.37	211.56	358.06	560.09
	L_∞	**810.77**	865.67	1616.80	927.25	1039.96	1094.87	1005.75	952.61

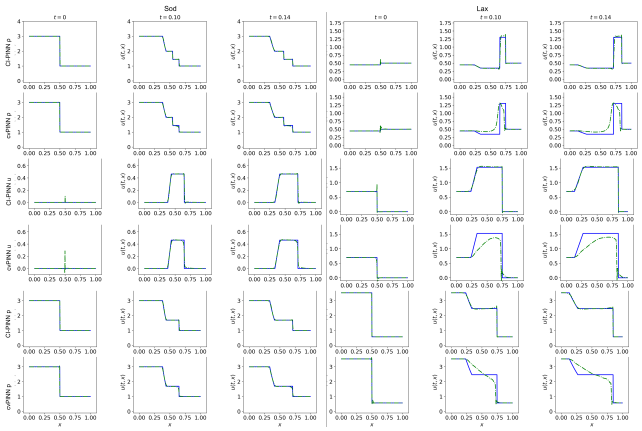


Figure 3. In these shock-tube profiles, the *contact discontinuity* appears as a sharp change in density while the velocity and pressure remain essentially unchanged across the interface. The *plateau states* are the flat segments between the smooth expansion region and the sharp jumps, and their constant levels are fixed by conservation across the discontinuities (Rankine–Hugoniot) together with the expansion relations.

This regime is challenging for standard PINNs and can be sensitive to flux non-convexity. CI-PINN resolves the sharp interface with the lowest shock-region L_1 error, without requiring auxiliary flux convexification, indicating robustness to more complex flux geometries.

Euler System (Sod & Lax Shock Tubes). For the 1D Euler system, CI-PINN and cvPINN consistently outperform the other baselines in Tables 3 and 4. Figure 3 visualizes representative profiles and highlights the key physical structures: a smooth expansion fan, a contact discontinuity (sharp density change with nearly unchanged (u, p)), a shock (jump in (ρ, u, p)), and the plateau states between them. In the Sod case, both methods recover the qualitative pattern, but cvPINN exhibits mild smearing around the contact and shock, leading to visibly perturbed plateau levels; The Lax case is substantially more challenging due to the stronger shock and steeper gradients: CI-PINN maintains sharp interfaces and accurate plateau states, indicating that the integral-form residual (enforcing Rankine–Hugoniot consistency) and the coupled consistency are effectively controlled even in the shock-dominated regime. In contrast, cvPINN often attains a reasonable shock location yet degrades the remaining structures—most notably the contact and plateau states—consistent with a smooth-surrogate solution that can reduce pointwise residuals but incurs a large accumulated mismatch in integral balance and admissibility near discontinuities. Overall, the results support the conclusion that integral-form enforcement is critical for correctly capturing shock/contact physics.

4.2. 2D Complex Flow Benchmarks

Compared to 1D, these 2D cases incur substantially higher training cost and do not admit a clean region-wise shock/rarefaction partition; we therefore report global space–time metrics and focus on comparisons against cvPINN as a strong baseline.

As shown in Figure 4, CI-PINN preserves the expected 2D dynamics. For SWE, the initial height hump develops into a coherent, radially propagating gravity-wave pattern. In contrast, cvPINN is visibly more diffusive: the wave ampli-

Table 4. Averaged error metrics (across u, p, ρ) for the Lax shock tube (values in 10^{-3}). Best is bold; second best is underlined. For full diagram please refer to appendix E

Region	Metric	CI-PINN	cvPINN	IPINN	Mod. MLP	Causal	RWF	AdaLR	Vanilla
Global	L_2	**55.89**	283.63	789.22	389.87	404.06	421.49	383.89	934.98
	L_1	**14.83**	142.00	600.61	245.62	237.60	249.23	286.91	828.75
	L_∞	1279.37	1630.57	2091.30	1443.45	**1301.50**	1519.52	1388.93	1738.78
Shock	L_2	**111.55**	492.82	1061.01	818.64	652.92	640.17	907.60	940.05
	L_1	**42.18**	282.83	646.66	497.16	451.16	525.91	451.15	938.51
	L_∞	1279.37	1630.57	2091.30	1443.45	**1301.50**	1519.52	1388.93	1736.84
Raref.	L_2	**168.17**	338.42	532.36	569.93	362.03	555.72	480.68	779.72
	L_1	**55.74**	268.49	309.37	488.70	266.65	555.12	443.96	852.34
	L_∞	1379.37	**1001.55**	2091.30	1504.43	1301.50	1519.53	1388.93	1738.77

Table 5. Global space-time error metrics (values in 10^{-3}). Top: 2D Euler Riemann problem. Bottom: 2D SWE. Best is **bold**; another is underlined.

Var	L_2		L_1		L_∞	
	CI-PINN	cvPINN	CI-PINN	cvPINN	CI-PINN	cvPINN
ρ	**82.46**	99.55	**25.76**	42.47	**1964.00**	2019.33
u	54.59	**53.29**	**14.13**	19.41	**997.63**	1235.11
v	47.64	**47.33**	**13.75**	20.85	874.52	**602.83**
p	**16.12**	29.50	**9.54**	17.66	**238.25**	496.03

(a) Euler 2D Riemann

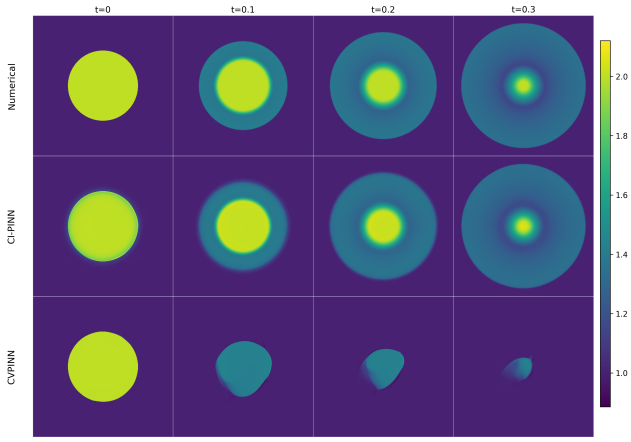
Var	L_2		L_1		L_∞	
	CI-PINN	cvPINN	CI-PINN	cvPINN	CI-PINN	cvPINN
h	**26.10**	225.81	**9.181**	127.875	**981.002**	1063.928
u	**16.37**	154.01	**5.365**	79.312	**223.043**	566.177
v	**16.25**	152.97	**5.522**	77.358	**185.416**	566.113

(b) SWE 2D

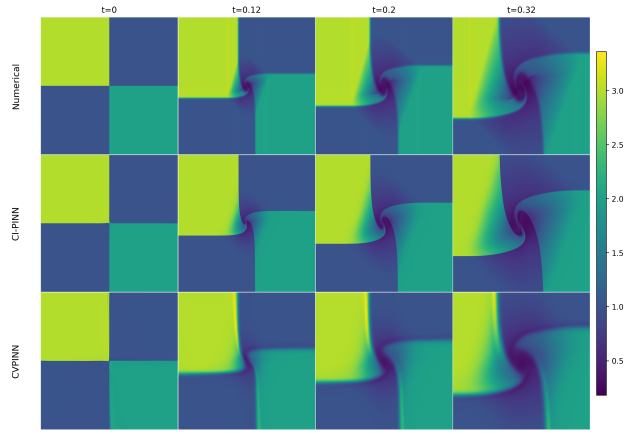
tude is strongly damped and the ring structure collapses over time. This failure is already apparent by $t \approx 0.2$, where the wavefront is largely smeared out, producing severe physical inaccuracies. CI-PINN, on the other hand, maintains circular symmetry and sustained outward propagation throughout the window $t \in [0, 0.3]$.

These qualitative differences are reflected quantitatively in Table 5(b). CI-PINN reduces the L_2 error by 88.4% for h ($225.81 \to 26.10$) and by 89.4% for (u, v) ($154.01 \to 16.37$; $152.97 \to 16.25$).

For the 2D Euler Riemann problem, CI-PINN more faithfully preserves sharp interfaces and the associated plateau states in the interacting-flow pattern (Figure 4), yielding clear gains in the key thermodynamic variables (Table 5(a)). In particular, ρ errors improve in L_1 by 39.3% ($42.47 \to 25.76$), with a 31.4% reduction in squared L_2 error (i.e., MSE) implied by $99.55 \to 82.46$. For pressure, the improvement is stronger: L_2 decreases by 45.3% ($29.50 \to 16.12$), L_1 by 46.0% ($17.66 \to 9.54$), and L_∞ by 52.0% ($496.03 \to 238.25$). Velocity errors are mixed in terms of L_2 but consistently better in L_1 (MAE reduced by 27.2% for u and 34.1% for v), while v-L_∞ can be dominated by a localized feature, leading to a larger maximum despite improved global fidelity.



(a) Height h in SWE



(b) Density ρ in Euler system

Figure 4. Selective component on 2D system, full result please refer to appendix J

5. Discussion and Limitations

We identified that strong-form PINNs fail at discontinuities because residuals diverge, creating an optimization barrier. We proposed the Coupled Integral PINN (CI-PINN) to circumvent this via dual-network integral conservation and entropy constraints. Evaluation on 1D and 2D benchmarks confirms that CI-PINN significantly outperforms the baselines considered, faithfully capturing shocks and contact discontinuities without mesh-based reconstruction.

Limitations. Although CI-PINN offers faster inference once trained, produces a continuous space-time solution that can be queried at arbitrary points without interpolation, and extends naturally to data assimilation and inverse problems, it does not match mature classical solvers on standard forward problems: on well-posed problems with simple geometry, finite-volume methods (e.g., MUSCL/WENO) are typically faster and more accurate. CI-PINN is therefore intended for the complementary regime where these advantages matter.

Impact Statement

This work advances physics-informed learning methods for hyperbolic PDEs by improving the reliability of neural PDE solvers in the presence of discontinuities (e.g., shocks). More accurate and stable surrogate solvers can reduce the cost of simulation and enable faster analysis in applications such as fluid dynamics, geophysics, and engineering design, which may have downstream benefits for safety and efficiency.

Potential risks mainly stem from misuse or over-trust: learned solvers can be applied outside their validity range (e.g., different regimes, boundary conditions, or resolutions) and may produce plausible-looking but incorrect solutions, which could misinform high-stakes decisions if used without verification. We mitigate this by emphasizing that the method is a numerical surrogate—not a guarantee of physical fidelity—and by encouraging standard validation against established baselines, reporting uncertainty/error metrics, and careful documentation of training domains and constraints. We do not anticipate direct negative societal impacts beyond the general concerns associated with applying machine-learning models to scientific and engineering workflows.

References

Bianchini, S. and Bressan, A. Vanishing viscosity solutions of nonlinear hyperbolic systems. *Annals of mathematics*, pp. 223–342, 2005.

Chaumet, A. and Giesselmann, J. Improving weak pinns for hyperbolic conservation laws: Dual norm computation, boundary conditions and systems, 2023.

Dafermos, C. M. and Dafermos, C. M. *Hyperbolic conservation laws in continuum physics*, volume 3. Springer, 2005.

Diab, W. and Kobaisi, M. A. Pinns for the solution of the hyperbolic buckley-leverett problem with a non-convex flux function, 2021. URL https://arxiv.org/abs/2112.14826.

Fuks, O. and Tchelepi, H. A. Limitations of physics informed machine learning for nonlinear two-phase transport in porous media. *Journal of Machine Learning for Modeling and Computing*, 1(1), 2020.

Goswami, S., Bora, A., Yu, Y., and Karniadakis, G. E. Physics-informed deep neural operator networks. In *Machine learning in modeling and simulation: methods and applications*, pp. 219–254. Springer, 2023.

Gupta, G., Xiao, X., and Bogdan, P. Multiwavelet-based operator learning for differential equations. *Advances in neural information processing systems*, 34:24048–24062, 2021.

Hansen, D., Maddix, D. C., Alizadeh, S., Gupta, G., and Mahoney, M. W. Learning physical models that can respect conservation laws. *Physica D: Nonlinear Phenomena*, 457:133952, 2024. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2023.133952. URL https://www.sciencedirect.com/science/article/pii/S0167278923003068.

Hopf, E. The partial differential equation ut + uux = μxx. *Communications on Pure and Applied Mathematics*, 3(3):201–230, 1950. doi: https://doi.org/10.1002/cpa.3160030302. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160030302.

Jagtap, A. D. and Karniadakis, G. E. Extended physics-informed neural networks (xpinns): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Communications in Computational Physics*, 28(5), 2020.

Jagtap, A. D., Kharazmi, E., and Karniadakis, G. E. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 365:113028, 2020.

Jeon, J. and Kim, S. J. Fvm network to reduce computational cost of cfd simulation. *arXiv preprint arXiv*, 2105, 2021.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Kharazmi, E., Zhang, Z., and Karniadakis, G. E. hp-vpinns: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021a. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2020.113547. URL https://www.sciencedirect.com/science/article/pii/S0045782520307325.

Kharazmi, E., Zhang, Z., and Karniadakis, G. E. hp-vpinns: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021b.

Krishnapriyan, A. S., Gholami, A., Zhe, S., Kirby, R. M., and Mahoney, M. W. Characterizing possible failure modes in physics-informed neural networks, 2021. URL https://arxiv.org/abs/2109.01050.

Lax, P. D. Weak solutions of nonlinear hyperbolic equations and their numerical computation. *Communications on Pure and Applied Mathematics*, 7(1):159–193, 1954. doi: https://doi.org/10.1002/cpa.3160070112. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160070112.

LeVeque, R. J. *The Derivation of Conservation Laws*, pp. 14–18. Birkhäuser Basel, Basel, 1992a. ISBN 978-3-0348-8629-1. doi: 10.1007/978-3-0348-8629-1_2. URL https://doi.org/10.1007/978-3-0348-8629-1_2.

LeVeque, R. J. *Rarefaction Waves and Integral Curves*, pp. 81–88. Birkhäuser Basel, Basel, 1992b. ISBN 978-3-0348-8629-1. doi: 10.1007/978-3-0348-8629-1_8. URL https://doi.org/10.1007/978-3-0348-8629-1_8.

LeVeque, R. J. *Shocks and the Hugoniot Locus*, pp. 70–80. Birkhäuser Basel, Basel, 1992c. ISBN 978-3-0348-8629-1. doi: 10.1007/978-3-0348-8629-1_7. URL https://doi.org/10.1007/978-3-0348-8629-1_7.

Li, T., Zhou, S., Chang, X., Zhang, L., and Deng, X. Finite volume graph network (fvgn): Predicting unsteady incompressible fluid dynamics with finite volume informed neural network. *arXiv preprint arXiv:2309.10050*, 2023.

Li, T., Zou, S., Chang, X., Zhang, L., and Deng, X. Predicting unsteady incompressible fluid dynamics with finite volume informed neural network. *Physics of Fluids*, 36 (4), 2024a.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., Azizzadenesheli, K., and Anandkumar, A. Physics-informed neural operator for learning partial differential equations. *ACM/JMS Journal of Data Science*, 1(3):1–27, 2024b.

Liu, L., Liu, S., Xie, H., Xiong, F., Yu, T., Xiao, M., Liu, L., and Yong, H. Discontinuity computing using physics-informed neural networks. *Journal of Scientific Computing*, 98(1):22, 2024.

Lu, L., Jin, P., and Karniadakis, G. E. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *CoRR*, abs/1910.03193, 2019. URL http://arxiv.org/abs/1910.03193.

Lu, L., Meng, X., Mao, Z., and Karniadakis, G. E. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, January 2021. ISSN 1095-7200. doi: 10.1137/19m1274067. URL http://dx.doi.org/10.1137/19M1274067.

Mao, Z., Jagtap, A. D., and Karniadakis, G. E. Physics-informed neural networks for high-speed flows. *Computer Methods in Applied Mechanics and Engineering*, 360:112789, 2020. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2019.112789. URL https://www.sciencedirect.com/science/article/pii/S0045782519306814.

Mei, D., Zhou, K., and Liu, C.-H. Unified finite-volume physics informed neural networks to solve the heterogeneous partial differential equations. *Knowledge-Based Systems*, 295:111831, 2024.

Muskat, M. The flow of homogeneous fluids through porous media. *Soil Science*, 46(2):169, 1938.

Patel, R. G., Manickam, I., Trask, N. A., Wood, M. A., Lee, M., Tomas, I., and Cyr, E. C. Thermodynamically consistent physics-informed neural networks for hyperbolic systems. *Journal of Computational Physics*, 449:110754, 2022. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2021.110754. URL https://www.sciencedirect.com/science/article/pii/S0021999121006495.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*, 2017.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Rajvanshi, M. P. and Ketcheson, D. I. Integral pinns for hyperbolic conservation laws. In *ICLR 2024 Workshop on AI4DifferentialEquations In Science*.

Ryck, T. D., Mishra, S., and Molinaro, R. wpinns: Weak physics informed neural networks for approximating entropy solutions of hyperbolic conservation laws, 2022.

Sod, G. A. A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws. *Journal of Computational Physics*, 27(1):1–31, 1978. ISSN 0021-9991. doi: https://doi.org/10.1016/0021-9991(78)90023-2. URL https://www.sciencedirect.com/science/article/pii/0021999178900232.

Souren Misra, S. V. R. R. and Bobba, M. K. Relaxation system based sub-grid scale modelling for large eddy simulation of burgers' equation. *International Journal of Computational Fluid Dynamics*, 24(8):303–315, 2010. doi: 10.1080/10618562.2010.523518. URL https://doi.org/10.1080/10618562.2010.523518.

Toro, E. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*. 01 2009. doi: 10.1007/b79761.

Toro, E. F. *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*. Springer Science & Business Media, 2013.

Wang, S., Teng, Y., and Perdikaris, P. Understanding and mitigating gradient pathologies in physics-informed neural networks, 2020. URL https://arxiv.org/abs/2001.04536.

Wang, S., Sankaran, S., and Perdikaris, P. Respecting causality is all you need for training physics-informed neural networks, 2022a.

Wang, S., Wang, H., Seidman, J. H., and Perdikaris, P. Random weight factorization improves the training of continuous neural representations, 2022b. URL https://arxiv.org/abs/2210.01274.

Wang, S., Sankaran, S., Wang, H., and Perdikaris, P. An expert's guide to training physics-informed neural networks, 2023.

Yu, B. et al. The deep ritz method: a deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.

Appendix

A. Proof sketch for propositions

Proposition(Strong-form blow-up as a discontinuity sharpens) Let $\mathbf{q}^*$ be a discontinuous physical solution containing a jump of amplitude $\Delta\mathbf{q} \neq \mathbf{0}$ across a smooth interface. Let $\{\mathbf{q}^\varepsilon\}_{\varepsilon>0}$ be a family of smooth approximations that resolves the jump over a transition layer of thickness ε. Assume $\mathbf{F} \in C^1$ and its Jacobians are bounded on the relevant state range. Assume additionally that the jump produces a nontrivial normal flux change, i.e. for a unit normal n to the interface,

$$\Delta\mathbf{F}_n := \mathbf{F}(\mathbf{q}_R)\cdot n - \mathbf{F}(\mathbf{q}_L)\cdot n \neq \mathbf{0}.$$

Then the strong-form residual satisfies

$$\mathcal{L}_{\text{strong}}(\mathbf{q}^\varepsilon) = \|\partial_t\mathbf{q}^\varepsilon + \nabla\cdot\mathbf{F}(\mathbf{q}^\varepsilon)\|^2_{L^2(\Omega_T)} \gtrsim \frac{C}{\varepsilon}, \qquad \varepsilon \to 0, \tag{6}$$

for some constant $C > 0$ independent of ε.

Proof (normal-line mean-value argument). Fix a time $t \in (0,T)$. Let $\Gamma(t) \subset \Omega$ denote the shock interface (a C^1 hypersurface), and pick a point $x_0 \in \Gamma(t)$. Let $n = n(x_0)$ be a unit normal at x_0. Assume the discontinuous (limit) solution has one-sided traces $\mathbf{q}_L := \lim_{\substack{x\to x_0\\(x-x_0)\cdot n<0}} \mathbf{q}^*(x,t)$, $\mathbf{q}_R := \lim_{\substack{x\to x_0\\(x-x_0)\cdot n>0}} \mathbf{q}^*(x,t)$, $\Delta\mathbf{q} := \mathbf{q}_R - \mathbf{q}_L \neq 0$. Let $\mathbf{q}^\varepsilon \in C^1(\Omega_T)$ be a smooth approximation that resolves this jump over a transition layer of thickness $\sim \varepsilon$ around $\Gamma(t)$. Concretely, along the normal line $x(s) := x_0 + s\,n$ with $|s| \lesssim \varepsilon$, the state transitions from a value near $\mathbf{q}_L$ to a value near $\mathbf{q}_R$.

Step 1: a $1/\varepsilon$ normal derivative appears (MVT). Define the 1D profile along the normal line

$$\phi(s) := \mathbf{q}^\varepsilon(x(s),t) \in \mathbb{R}^m.$$

By the transition-layer assumption, there exist $s_- < s_+$ with $|s_\pm| \lesssim \varepsilon$ such that $\phi(s_-) \approx \mathbf{q}_L$, $\phi(s_+) \approx \mathbf{q}_R$, hence $\|\phi(s_+) - \phi(s_-)\| \sim \|\Delta\mathbf{q}\|$. Since ϕ is C^1, the (1D) mean value theorem applied componentwise gives a point $\xi \in (s_-, s_+)$ with

$$\|\phi'(\xi)\| \gtrsim \frac{\|\phi(s_+) - \phi(s_-)\|}{|s_+ - s_-|} \sim \frac{\|\Delta\mathbf{q}\|}{\varepsilon} \sim \frac{1}{\varepsilon}.$$

But $\phi'(s) = \nabla\mathbf{q}^\varepsilon(x(s),t)\,n$, i.e. the directional derivative along the normal. Thus at some point in the layer,

$$\|\partial_n\mathbf{q}^\varepsilon\| := \|\nabla\mathbf{q}^\varepsilon\, n\| \sim \frac{1}{\varepsilon}.$$

Step 2: why this forces $\nabla\cdot\mathbf{F}(\mathbf{q}^\varepsilon) \sim 1/\varepsilon$. Write the flux as $\mathbf{F}(\mathbf{q}) = (\mathbf{F}_1(\mathbf{q}),\ldots,\mathbf{F}_d(\mathbf{q}))$ with $\mathbf{F}_j(\mathbf{q}) \in \mathbb{R}^m$. Introduce the *normal flux vector* $\mathbf{F}_n(\mathbf{q}) := \sum_{j=1}^d n_j\,\mathbf{F}_j(\mathbf{q}) \in \mathbb{R}^m$, and its Jacobian $\mathbf{J}_{F_n}(\mathbf{q}) = \sum_{j=1}^d n_j\,\mathbf{J}_{\mathbf{F}_j}(\mathbf{q})$.

A standard local-coordinate decomposition (normal n + tangential directions $\tau_1,\ldots,\tau_{d-1}$) gives, near x_0,

$$\nabla\cdot\mathbf{F}(\mathbf{q}^\varepsilon) = \partial_n\mathbf{F}_n(\mathbf{q}^\varepsilon) + \sum_{k=1}^{d-1}\partial_{\tau_k}\mathbf{F}_{\tau_k}(\mathbf{q}^\varepsilon) + (\text{curvature terms}), \tag{7}$$

where $\mathbf{F}_{\tau_k}(\mathbf{q}) := \sum_{j=1}^d (\tau_k)_j\,\mathbf{F}_j(\mathbf{q})$. The key point is: *only the ∂_n term differentiates across the thin layer*. By assumption, $\mathbf{q}^\varepsilon$ varies sharply only in the normal direction (thickness $\sim \varepsilon$) and remains $O(1)$-smooth tangentially; hence the tangential/curvature terms are $O(1)$, while $\partial_n\mathbf{F}_n(\mathbf{q}^\varepsilon) = \mathbf{J}_{F_n}(\mathbf{q}^\varepsilon)\,\partial_n\mathbf{q}^\varepsilon$ (chain rule in the normal direction). Since the Jacobians are bounded and the jump in normal flux is $O(1)$ whenever $\Delta\mathbf{q} \neq 0$, one generically has $\|\mathbf{J}_{F_n}(\mathbf{q}^\varepsilon)\| = O(1)$ in the layer and thus

$$\|\partial_n\mathbf{F}_n(\mathbf{q}^\varepsilon)\| \sim \|\partial_n\mathbf{q}^\varepsilon\| \sim \frac{1}{\varepsilon}.$$

Plugging into (7) yields

$$\|\nabla\cdot\mathbf{F}(\mathbf{q}^\varepsilon)\| \sim \frac{1}{\varepsilon} \quad \text{on (part of) the layer.}$$

Step 3: integrate over a set of space–time measure $O(\varepsilon)$. The transition layer is an ε-thick band around a codimension-1 interface, so its space–time measure scales like $O(\varepsilon)$. On this band, the residual $\mathbf{r}^\varepsilon := \partial_t \mathbf{q}^\varepsilon + \nabla \cdot \mathbf{F}(\mathbf{q}^\varepsilon)$ satisfies

$$\|\mathbf{r}^\varepsilon\| \;\geq\; \|\nabla \cdot \mathbf{F}(\mathbf{q}^\varepsilon)\| - \|\partial_t \mathbf{q}^\varepsilon\|.$$

If $\|\partial_t \mathbf{q}^\varepsilon\| = O(1)$ in the layer, it is dominated by the $O(1/\varepsilon)$ divergence term as $\varepsilon \to 0$; if $\|\partial_t \mathbf{q}^\varepsilon\|$ also scales like $1/\varepsilon$, then $\|\mathbf{r}^\varepsilon\| \sim 1/\varepsilon$ in magnitude (up to non-generic cancellations). Therefore $\|\mathbf{r}^\varepsilon\| \sim 1/\varepsilon$ on a set of measure $O(\varepsilon)$, and hence

$$\mathcal{L}_{\text{strong}}(\mathbf{q}^\varepsilon) = \int_{\Omega_T} \|\mathbf{r}^\varepsilon\|^2 \, dx \, dt \;\gtrsim\; O(\varepsilon) \cdot \left(\frac{1}{\varepsilon}\right)^2 \;=\; \frac{C}{\varepsilon},$$

for some $C > 0$ independent of ε which conclude the proof for proposition 2.3 and 2.1.

□

B. Benchmarks

We compare CI-PINN against several advances methods in physics-informed neural networks. **Adaptive learning rate (Adaptive lr)** (Wang et al., 2023) dynamically adjusts optimizer step sizes based on gradient magnitudes, ensuring balanced contributions from each loss component and improving convergence stability. **Causal training (Causal)** (Wang et al., 2022a) introduces a temporally structured curriculum that gradually incorporates the residual loss to align with the causal nature of time-dependent PDEs, thereby enhancing long-horizon stability. **Modified MLP** (Wang et al., 2020) increases model expressiveness by processing inputs through parallel subnetworks and aggregating outputs via gated mechanisms, significantly boosting accuracy with minimal computational overhead. **Random Weight Factorization (RWF)** (Wang et al., 2022b) reparameterizes neural network weights as products of random matrices, mitigating spectral bias and smoothing the optimization landscape in PINN training.

Among conservation-based formulations, **cvPINN** (Patel et al., 2022) integrates conservation laws over local control volumes using numerical quadrature, inspired by finite-volume methods, to enforce local conservation properties. In contrast, **IPINN** (Rajvanshi & Ketcheson) directly approximates the spatial integral of the solution using neural networks, enforcing global conservation without explicit discretization. These integral-form methods provide more robust handling of shocks and discontinuities than standard differential-form PINNs. Additionally, we include a small-scale comparison with the **ProbConserv** Bayesian framework (Hansen et al., 2024), which imposes global conservation via probabilistic posterior updates on the predicted solution. Although this method requires problem-specific boundary treatment and is less generalizable to complex settings. We report the formulations and results of **cvPINN** and **IPINN** in the main text, while the performance of all other mentioned methods is provided in Appendix H as a point of reference.

C. Rigorous derivation of the Integral Form of Scalar Conservation Law

In nonlinear conservation law, the shocks appear when the characteristic lines of different values intersect with each other (LeVeque, 1992a). The properties of these shocks are governed by the weak form of the PDE (i.e. Rankine-Hugoniot jump condition). The original conservation law can be written as:

$$u_t + f(u)_x = 0 \tag{8}$$

where u is conserved quantity, u_t denotes the partial derivative of u with respect to time t, $f(u)$ is is a flux function that describes how the quantity u flows or moves through space, $f(u)_x$ is denotes the partial derivative of $f(u)$ with respect to space.

For the weak form of a conservation law, we multiply the equation by a test function $\phi : \mathbb{R} \to \mathbb{R}$ and integrate over space and time:

$$\int_0^\infty \int_{-\infty}^{+\infty} [\phi u_t + \phi f_x] \, dx \, dt = 0 \tag{9}$$

Due to the conservation form of the PDE, this equation must hold for any test function ϕ. To derive the integral form, we select a specific test function ϕ defined as:

$$\phi(x,t) = \begin{cases} 1 & \text{for } (x,t) \in [x_1, x_2] \times [t_1, t_2] \\ 0 & \text{for } (x,t) \notin [x_1 - \epsilon, x_2 + \epsilon] \times [t_1 - \epsilon, t_2 + \epsilon] \end{cases} \tag{10}$$

with ϕ smooth in the transition layer of width ϵ. Letting $\epsilon \to 0$ yields the conservative integral balance (LeVeque, 1992a)

$$\begin{aligned} &\int_{x_1}^{x_2} u(x,t_2)\,dx - \int_{x_1}^{x_2} u(x,t_1)\,dx \\ &+ \int_{t_1}^{t_2} \Big(f(u(x_2,t)) - f(u(x_1,t)) \Big)\,dt = 0, \end{aligned} \tag{11}$$

or, equivalently,

$$\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)\,dx = f(u(x_1,t)) - f(u(x_2,t)). \tag{12}$$

which holds for all x_1, x_2, and t. This integral form is the basis for Finite Volume Methods (FVM). For smooth solutions, differentiating the integral balance with respect to the control volume recovers the differential form. For discontinuous solutions, the integral/weak form remains meaningful while the strong form does not. This method can also be extended to vector space later

Rankine–Hugoniot jump condition (shock speed). Suppose u has a single shock located at $x = s(t)$ that separates left and right states u_L and u_R. Consider a thin space-time control volume that straddles the shock curve. Applying the conservation balance (11) and letting the thickness shrink to zero yields the Rankine–Hugoniot condition:

$$\dot{s}(t)\,[u] = [f(u)],\ [u] := u_R - u_L,\ [f(u)] := f(u_R) - f(u_L), \tag{13}$$

so that the shock speed is

$$\dot{s}(t) = \frac{f(u_R) - f(u_L)}{u_R - u_L}. \tag{14}$$

This jump condition is the statement that the discontinuity moves in a way that preserves the integral conservation law.

To extend to vector form, given in Equation (15):

$$\int_\Omega u(t,x)\,\mathrm{d}\Omega = \int_\Omega h(x)\,\mathrm{d}\Omega - \int_0^t \int_\Gamma F(u)\cdot n\,\mathrm{d}\Gamma\,\mathrm{d}t, \tag{15}$$

we first integrate the differential form of the conservation law, given in Equation (16):

$$\mathcal{F}u \;=\; u_t \;+\; \nabla\cdot F(u), \tag{16}$$

over the spatial domain Ω. From this, we obtain an expression for the rate of change in time of the total conserved quantity:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u(t,x)\,\mathrm{d}\Omega = \int_\Omega u_t(t,x)\,\mathrm{d}\Omega, \tag{17a}$$

$$= -\int_\Omega \nabla\cdot F(u)\,\mathrm{d}\Omega, \tag{17b}$$

$$= -\int_\Gamma F(u)\cdot n\,\mathrm{d}\Gamma. \tag{17c}$$

Next, integrating Equation (17a) in time from 0 to t gives

$$\int_0^t \int_\Omega u_\tau(\tau,x)\,\mathrm{d}\Omega\,\mathrm{d}\tau = \int_\Omega u(t,x)\,\mathrm{d}\Omega - \int_\Omega u(0,x)\,\mathrm{d}\Omega, \tag{18}$$

where $u(0,x) = h(x)$ is the prescribed initial condition. Equating the right-hand side of (18) with the temporal integral of (17c) recovers the integral form (15).

D. Mechanism of Shocks and Rarefactions

We consider a nonlinear hyperbolic conservation law in $\Omega \subset \mathbb{R}^d$,

$$\partial_t \mathbf{q}(\mathbf{x}, t) \;+\; \nabla \cdot \mathbf{F}\big(\mathbf{q}(\mathbf{x}, t)\big) \;=\; \mathbf{0}, \tag{19}$$

where $\mathbf{q}(\mathbf{x}, t) \in \mathbb{R}^m$ denotes the conserved variables and $\mathbf{F}(\mathbf{q}) = [\mathbf{F}_1(\mathbf{q}), \ldots, \mathbf{F}_d(\mathbf{q})]$ is the flux with $\mathbf{F}_j(\mathbf{q}) \in \mathbb{R}^m$. Hyperbolicity means that, for any unit direction $\mathbf{n} \in \mathbb{S}^{d-1}$, the directional flux Jacobian

$$\mathbf{A}(\mathbf{q}; \mathbf{n}) \;:=\; \sum_{j=1}^{d} n_j \, \nabla_{\mathbf{q}} \mathbf{F}_j(\mathbf{q}) \;\in\; \mathbb{R}^{m \times m} \tag{20}$$

is real diagonalizable, admitting eigenpairs $\{(\lambda_i(\mathbf{q}; \mathbf{n}), \mathbf{r}_i(\mathbf{q}; \mathbf{n}))\}_{i=1}^m$.

To describe the local wave pattern, consider the planar Riemann initial data in direction $\mathbf{n}$,

$$\mathbf{q}(\mathbf{x}, 0) = \begin{cases} \mathbf{q}_L, & \mathbf{n} \cdot \mathbf{x} < 0, \\ \mathbf{q}_R, & \mathbf{n} \cdot \mathbf{x} > 0. \end{cases} \tag{21}$$

For each characteristic family $i \in \{1, \ldots, m\}$, the solution consists of elementary waves determined by the ordering of the characteristic speeds across the jump.

- **Shock waves (compressive).** If $\lambda_i(\mathbf{q}_L; \mathbf{n}) > \lambda_i(\mathbf{q}_R; \mathbf{n})$, then i-characteristics enter the discontinuity and the jump persists as a shock. The shock propagates with speed s along $\mathbf{n}$ and satisfies the *Rankine–Hugoniot* condition

$$s \, [\![\mathbf{q}]\!] \;=\; [\![\mathbf{F}(\mathbf{q})]\!]\mathbf{n} \;=\; \sum_{j=1}^{d} n_j \, [\![\mathbf{F}_j(\mathbf{q})]\!], \qquad [\![\mathbf{q}]\!] := \mathbf{q}_R - \mathbf{q}_L, \tag{22}$$

  together with the *Lax entropy condition*

$$\lambda_i(\mathbf{q}_R; \mathbf{n}) \;<\; s \;<\; \lambda_i(\mathbf{q}_L; \mathbf{n}), \tag{23}$$

  which selects the physically admissible (compressive) shock.

- **Rarefaction waves (expansive).** If $\lambda_i(\mathbf{q}_L; \mathbf{n}) < \lambda_i(\mathbf{q}_R; \mathbf{n})$, then i-characteristics diverge and the discontinuity spreads into a continuous rarefaction fan. The rarefaction admits a self-similar form $\mathbf{q}(\mathbf{x}, t) = \mathbf{\Phi}(\xi)$ with similarity variable $\xi = \frac{\mathbf{n} \cdot \mathbf{x}}{t}$. In smooth regions, $\mathbf{\Phi}$ satisfies the ODE

$$\frac{d\mathbf{\Phi}}{d\xi} \;\parallel\; \mathbf{r}_i\big(\mathbf{\Phi}; \mathbf{n}\big), \qquad \xi = \lambda_i\big(\mathbf{\Phi}; \mathbf{n}\big), \tag{24}$$

  with boundary states $\mathbf{\Phi}(\lambda_i(\mathbf{q}_L; \mathbf{n})) = \mathbf{q}_L$ and $\mathbf{\Phi}(\lambda_i(\mathbf{q}_R; \mathbf{n})) = \mathbf{q}_R$.

The above characterization explains how discontinuities and smooth fans emerge from the same conservation structure: shocks correspond to compressive wave interactions, while rarefactions correspond to expansive wave spreading. A detailed treatment can be found in standard references on hyperbolic conservation laws (LeVeque, 1992c;b).

E. Full Table of Euler system

Table A1. Error metrics for Sod and Lax shock tubes (values shown in 10^{-3} units). Each entry reports (Normalized L_2, Normalized L_1, L_∞). Best is **bold**; second best is underlined.

Case	Comp.	Region	Metric	CI-PINN	cvPINN	IPINN	Modified_MLP	Causal	RWF	Adaptive lr	Vanilla PINN
Sod	u	Global	Norm. L_2	**20.68**	20.94	243.46	206.73	226.93	209.91	149.47	243.91
			Norm. L_1	**4.42**	6.86	138.43	147.04	126.42	147.20	113.01	133.38
			L_∞	412.73	443.43	689.79	695.54	638.42	790.95	**407.81**	464.31
		Shock	Norm. L_2	45.02	**44.92**	316.16	361.16	218.41	376.38	212.58	314.88
			Norm. L_1	**12.99**	22.15	218.10	313.58	155.21	326.31	189.30	214.45
			L_∞	412.73	443.43	689.79	695.54	602.79	790.95	**407.81**	464.31
		Rarefaction	Norm. L_2	57.97	**53.44**	275.83	150.17	96.94	147.68	195.19	274.90
			Norm. L_1	**24.80**	26.03	232.65	110.49	76.96	160.69	163.70	232.01
			L_∞	412.73	443.43	559.87	695.54	418.82	790.95	**407.81**	464.31
	p	Global	Norm. L_2	**30.75**	37.32	435.16	214.04	388.12	205.84	387.37	508.67
			Norm. L_1	**8.24**	14.78	268.63	152.58	223.48	141.56	258.91	270.93
			L_∞	1213.53	1517.11	6311.43	**995.76**	1297.68	1016.46	1451.26	1322.80
		Shock	Norm. L_2	**61.12**	74.45	462.70	343.41	540.51	340.50	457.73	502.94
			Norm. L_1	**18.37**	38.80	334.03	288.09	382.49	280.34	360.26	336.94
			L_∞	1213.53	1517.11	5158.10	**995.76**	1297.68	1016.46	1451.26	1322.65
		Rarefaction	Norm. L_2	**92.06**	94.89	358.70	282.70	297.40	273.91	585.86	795.28
			Norm. L_1	**55.51**	58.92	296.01	237.44	248.81	229.05	482.13	685.69
			L_∞	806.56	**719.72**	956.57	995.76	1144.46	1016.46	1451.26	1322.63
	ρ	Global	Norm. L_2	**26.40**	37.87	996.53	183.28	492.86	185.19	228.64	425.07
			Norm. L_1	**7.64**	13.89	673.78	131.72	259.88	132.21	169.88	222.89
			L_∞	1213.02	1433.86	6135.32	1090.46	1696.85	1477.19	1158.17	**1070.88**
		Shock	Norm. L_2	**43.38**	63.85	1315.96	268.77	812.70	273.72	292.70	579.14
			Norm. L_1	**14.71**	31.41	1050.42	229.84	624.69	237.73	253.67	432.20
			L_∞	1213.02	1433.86	6135.32	1090.46	1696.85	1477.19	1158.17	**1070.88**
		Rarefaction	Norm. L_2	**80.32**	93.87	1507.81	208.97	398.77	213.09	293.12	610.09
			Norm. L_1	**48.48**	54.09	1451.58	167.36	325.41	163.99	245.43	524.54
			L_∞	1213.02	1433.86	3333.95	1090.46	1556.61	1477.19	1158.17	**1070.88**
Lax	u	Global	Norm. L_2	**44.13**	282.17	629.87	410.98	298.65	443.50	453.70	887.05
			Norm. L_1	**16.28**	158.04	496.63	285.26	149.48	282.50	372.99	747.72
			L_∞	1157.08	1736.74	1385.39	1352.67	1326.55	1523.93	**1084.90**	1318.13
		Shock	Norm. L_2	**83.66**	452.16	776.69	644.51	568.98	719.75	610.56	930.47
			Norm. L_1	**22.13**	255.35	694.10	553.33	360.00	580.34	564.90	752.99
			L_∞	1157.08	1736.74	1385.39	1352.67	1326.55	1523.93	**1084.90**	1318.13
		Rarefaction	Norm. L_2	**124.42**	390.88	552.93	625.57	349.88	715.72	559.44	920.95
			Norm. L_1	**57.73**	319.74	486.86	562.44	264.66	633.13	541.40	879.62
			L_∞	1457.08	1736.74	1385.39	1535.63	1326.55	1523.94	**1084.90**	1318.11
	p	Global	Norm. L_2	**70.24**	357.60	689.18	554.36	642.40	605.80	493.73	1054.43
			Norm. L_1	**20.09**	177.78	495.92	349.94	401.85	354.85	392.65	929.45
			L_∞	1875.14	2184.09	**1703.83**	1908.40	1809.39	1900.42	2186.46	2211.95
		Shock	Norm. L_2	**146.60**	643.03	1090.38	878.63	959.83	956.80	636.79	1310.55
			Norm. L_1	**40.79**	355.99	837.22	682.19	651.30	718.03	522.61	988.26
			L_∞	1875.14	2184.09	**1703.83**	1908.40	1809.39	1900.42	2186.46	2206.13
		Rarefaction	Norm. L_2	**292.06**	496.41	436.28	875.24	522.34	732.43	737.43	1145.92
			Norm. L_1	**78.35**	405.21	336.09	793.96	426.60	908.75	709.77	996.15
			L_∞	1875.14	**297.01**	1703.83	1908.40	1809.39	1900.42	2186.46	2211.95
	ρ	Global	Norm. L_2	**53.29**	211.11	1048.62	204.26	271.12	215.17	204.25	863.43
			Norm. L_1	**8.12**	90.19	809.30	101.65	161.48	110.34	95.09	809.09
			L_∞	805.88	970.89	3184.67	1069.27	**768.57**	1134.22	895.44	1686.26
		Shock	Norm. L_2	**104.38**	383.28	1315.96	932.79	429.95	243.97	1475.46	579.13
			Norm. L_1	**63.61**	237.16	408.67	255.95	342.17	279.35	265.95	1074.29
			L_∞	805.88	970.89	3184.67	1069.27	**768.57**	1134.22	895.44	1686.26
		Rarefaction	Norm. L_2	**88.03**	127.97	607.86	208.97	213.86	219.02	145.16	272.28
			Norm. L_1	**31.13**	80.51	105.17	109.69	108.68	123.48	80.71	681.26
			L_∞	805.88	970.89	3184.67	1069.27	**768.57**	1134.22	895.44	1686.26

F. Clips plots

F.1. Burgers Equation

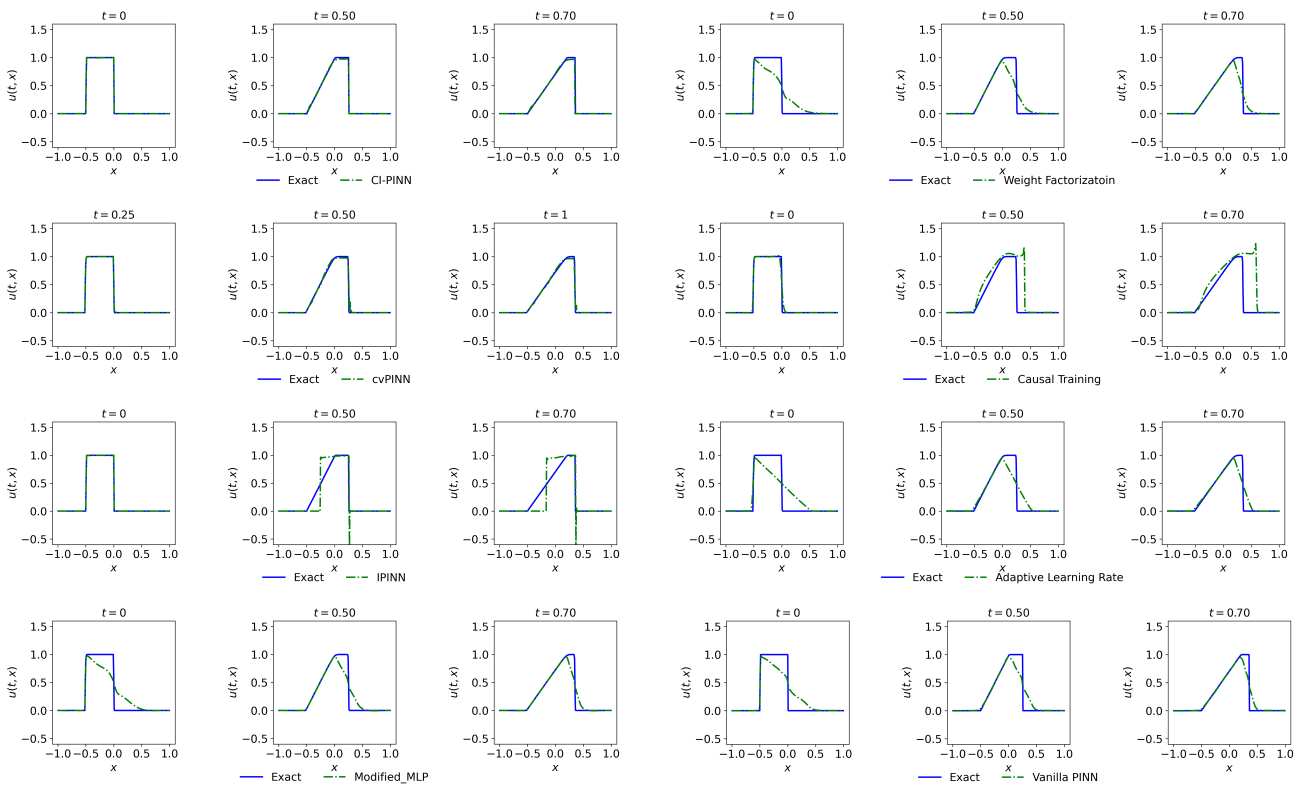


Figure A1. Comparison of predictions for Burger problem.

F.2. BL Equation

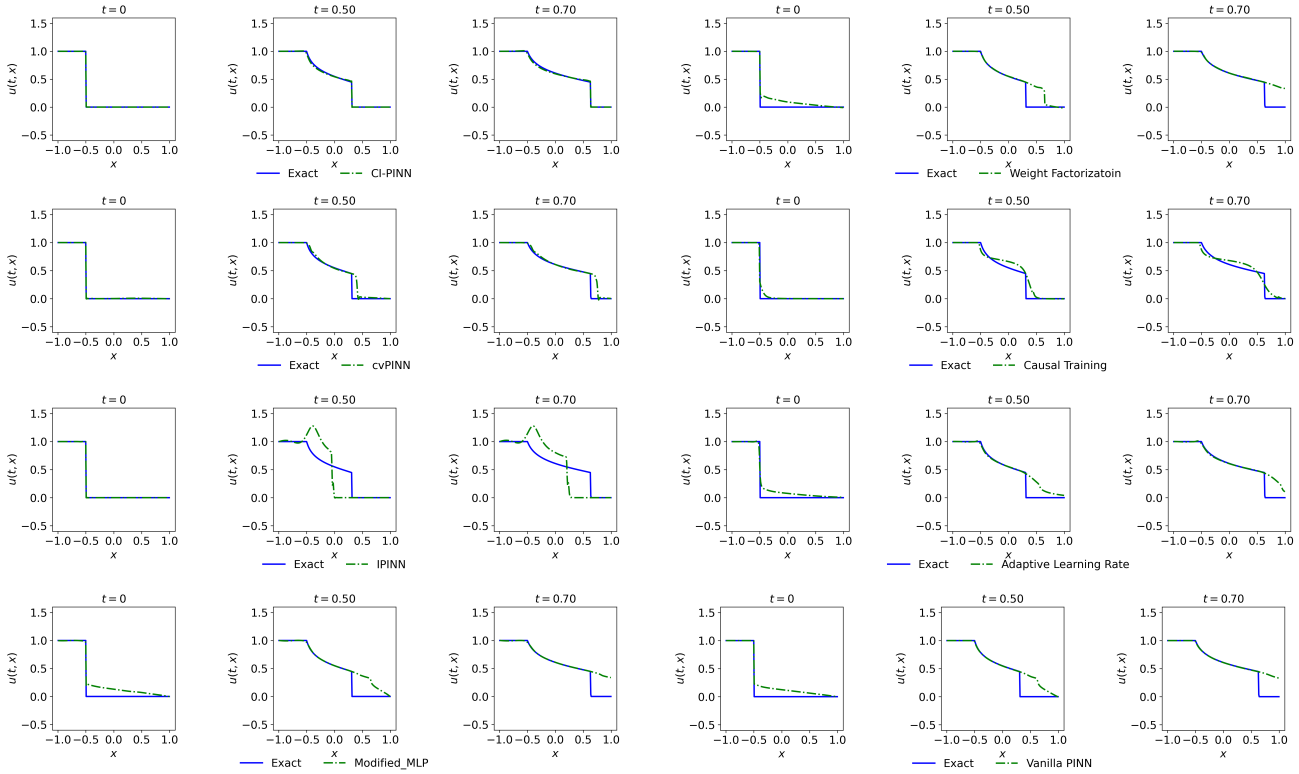


Figure A2. Comparison of predictions for BL problem.

F.3. Sod Problem

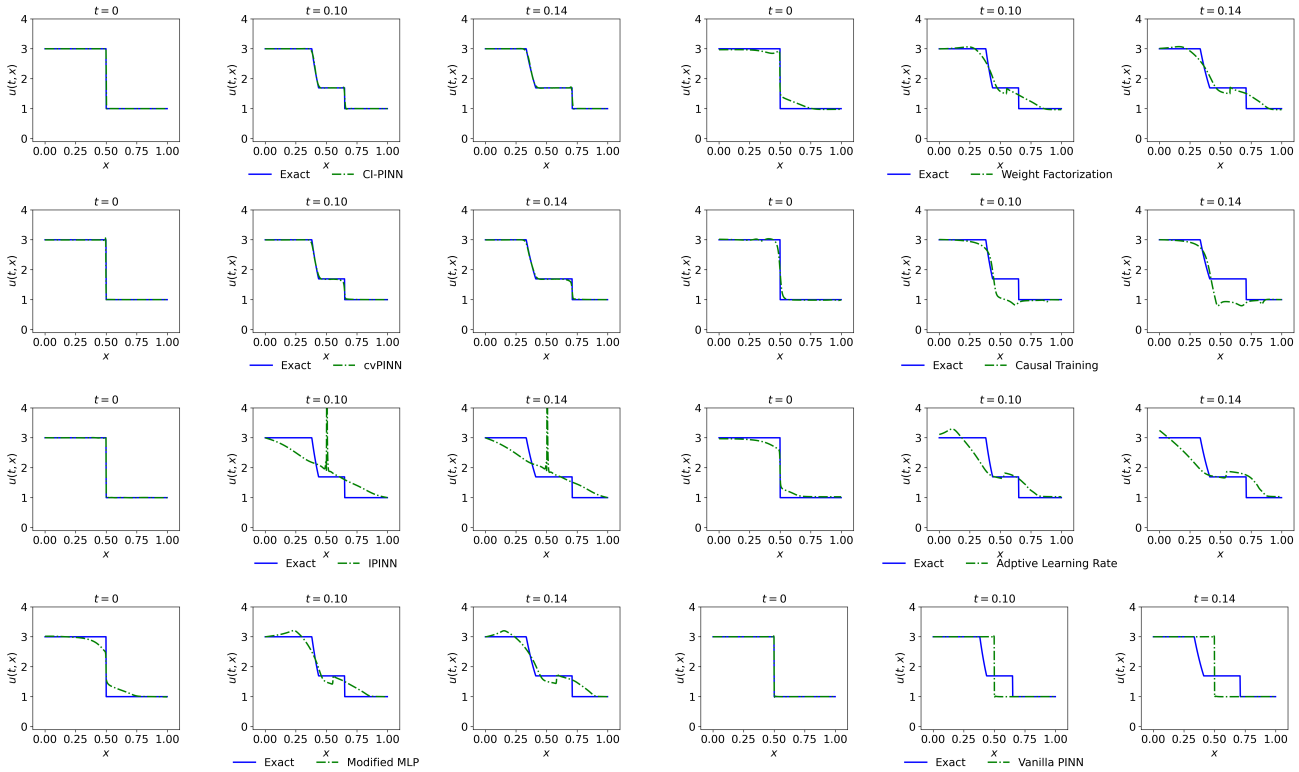


Figure A3. Sod problem: p component.

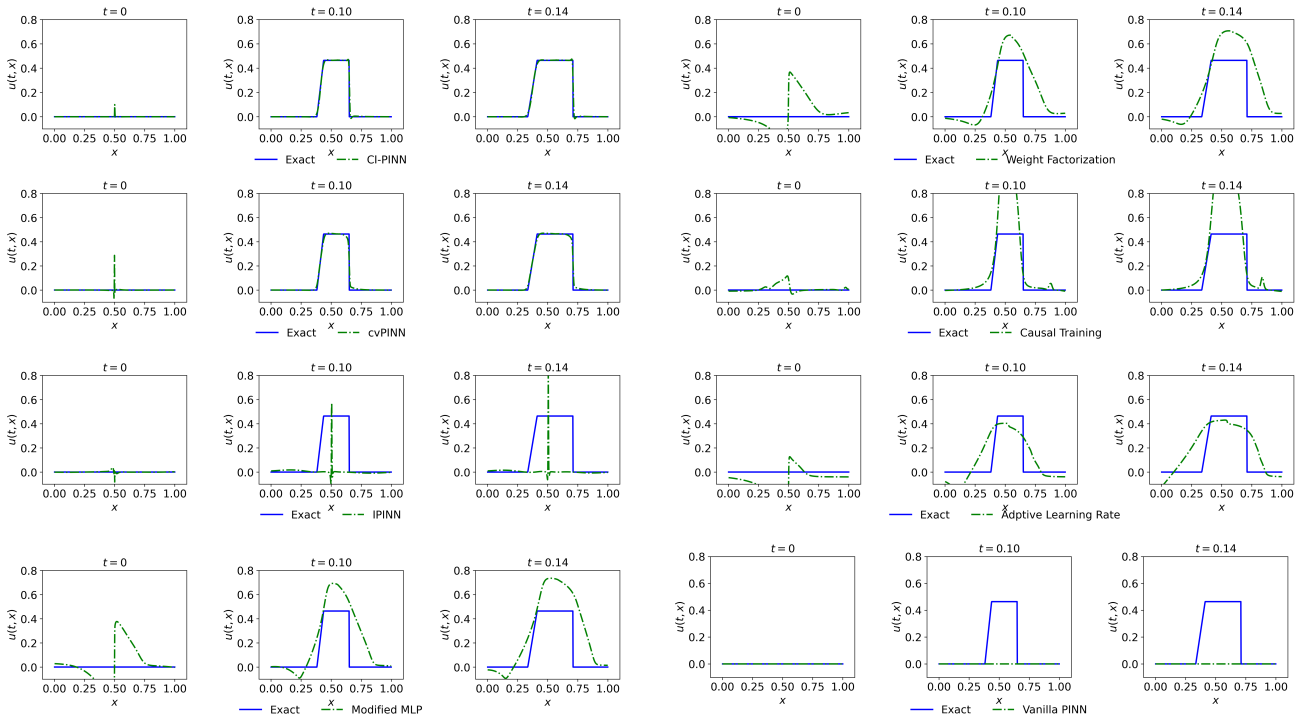


Figure A4. Sod problem: u component.

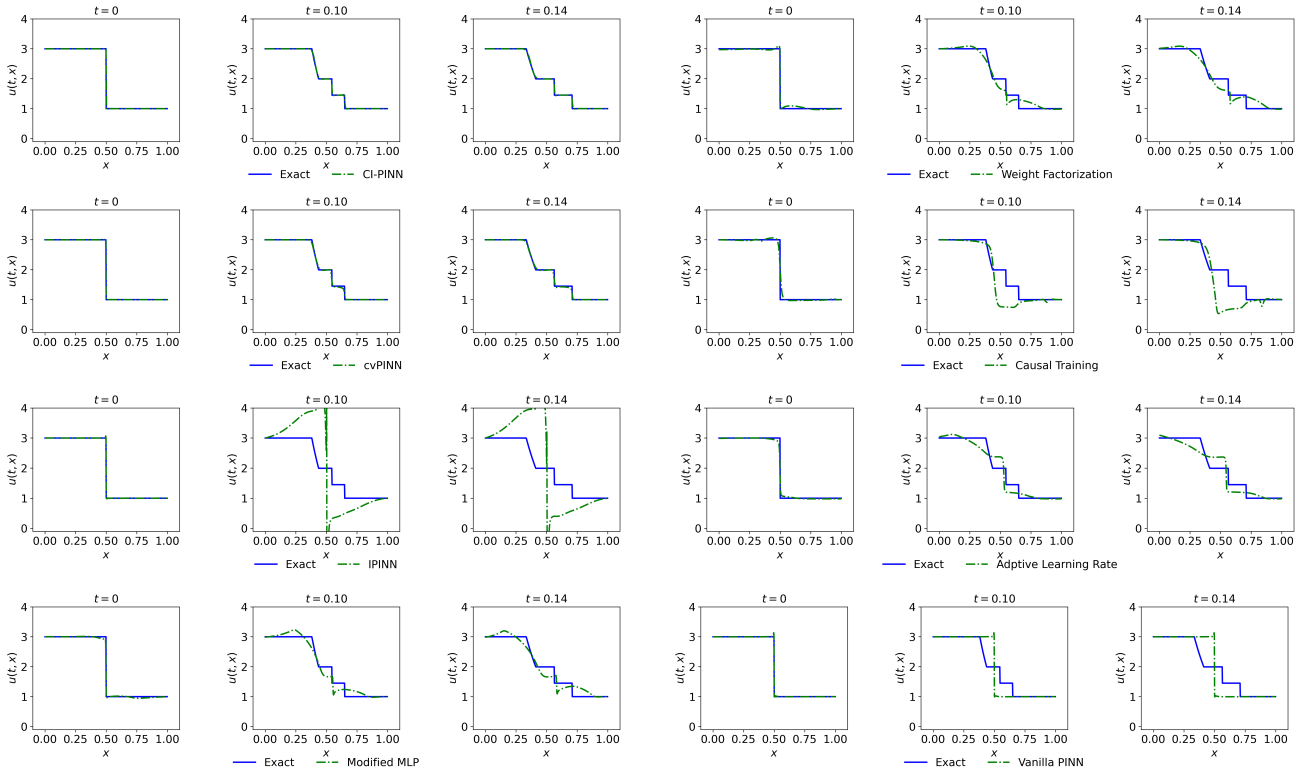


Figure A5. Sod problem: ρ component.

F.4. Lax Problem

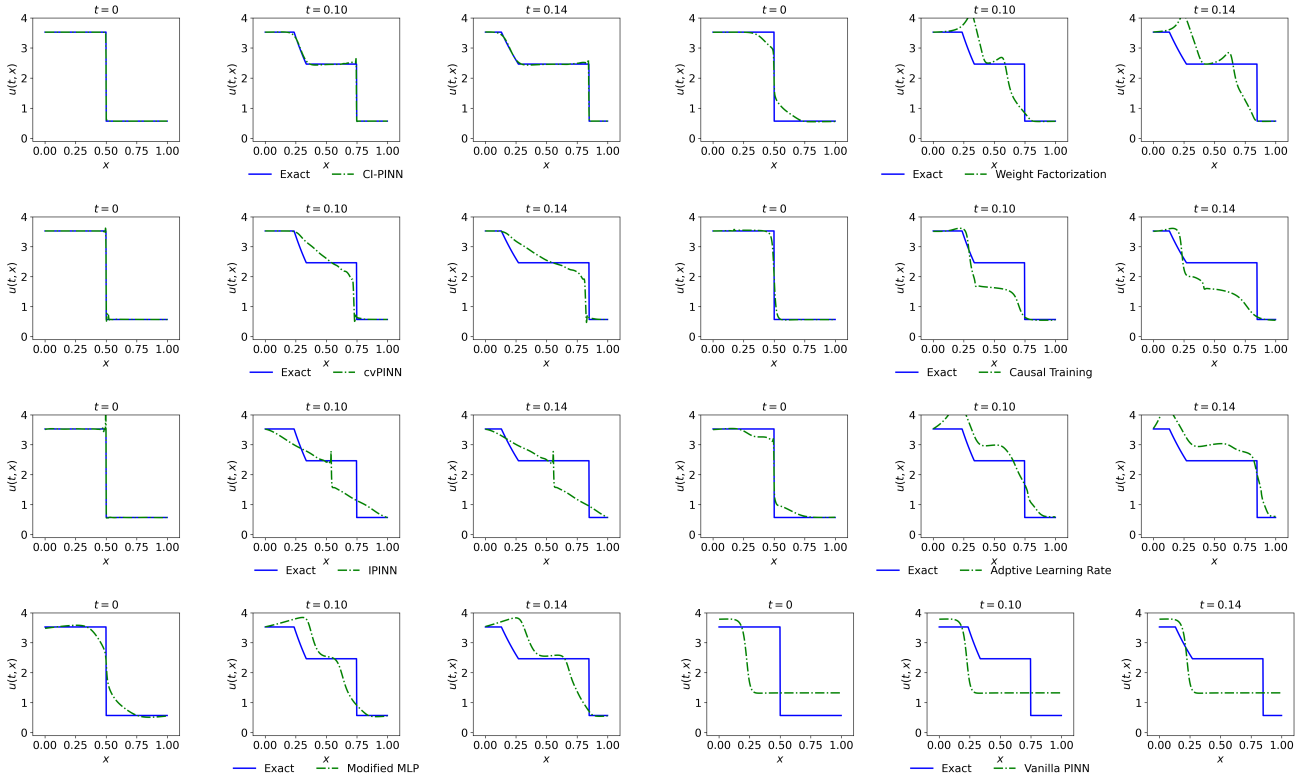


Figure A6. Lax problem: p component.

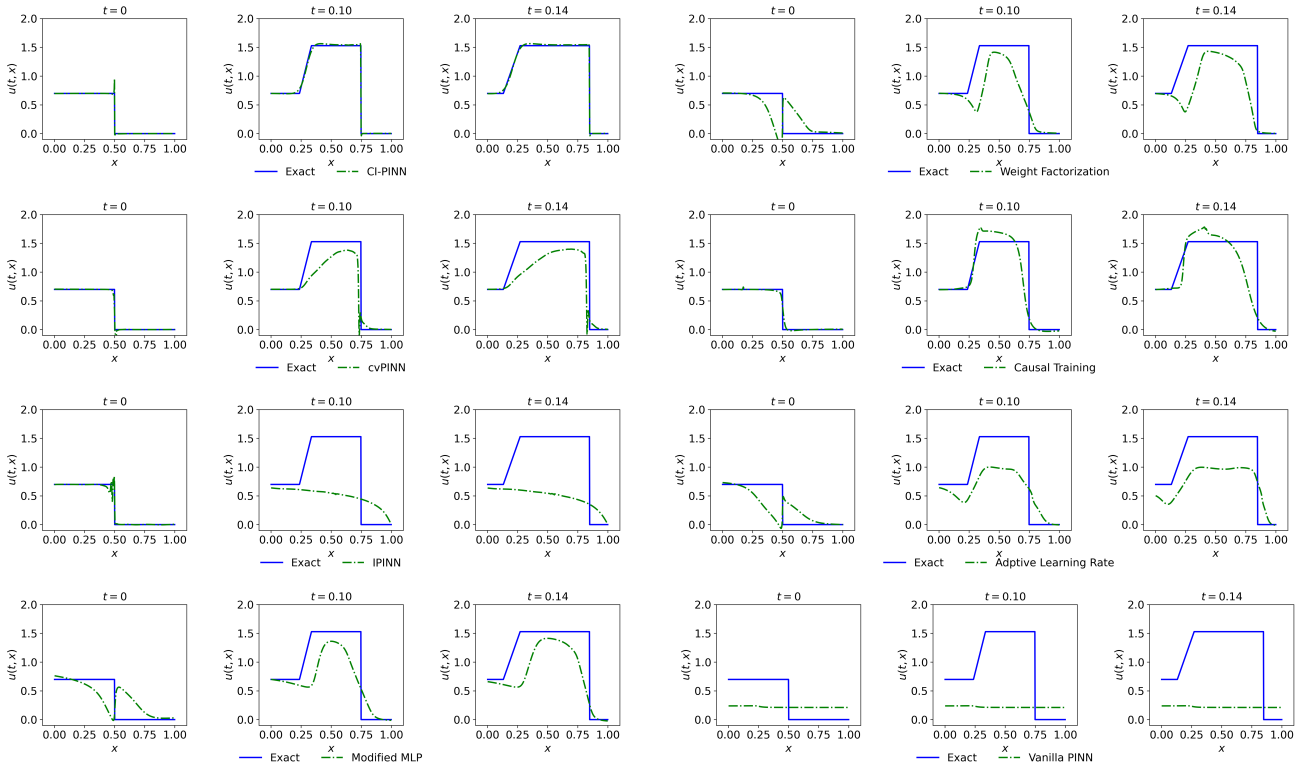


Figure A7. Lax problem: u component.

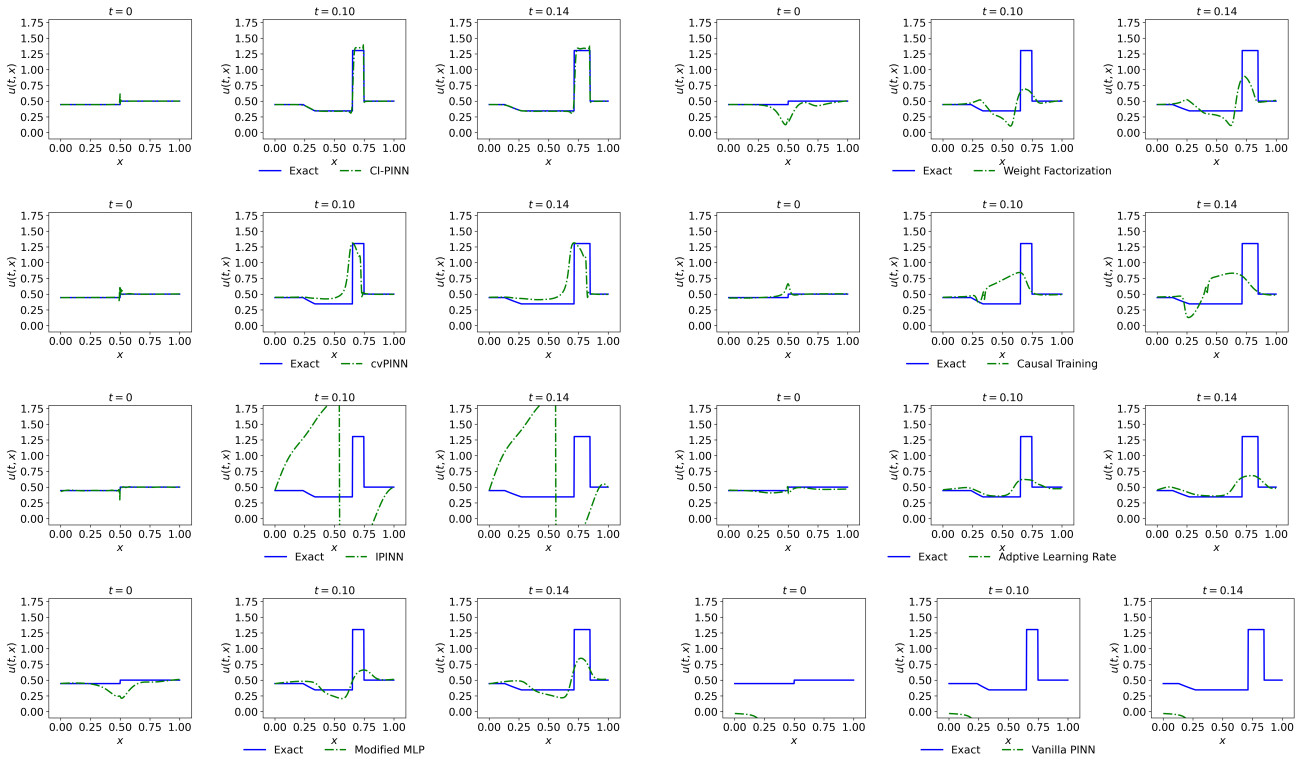


Figure A8. Lax problem: ρ component.

G. Domain-Partition

We split the space–time domain $\Omega \times [0, T]$, with $\Omega \subset \mathbb{R}^d$, into three parts: the shock neighborhood Ω_s, the rarefaction region Ω_r, and the full domain Ω. See appendix for detailed definition.

Let $\mathbf{x} \in \mathbb{R}^d$ be the spatial coordinate and define the spatial diameter

$$W \;=\; \mathrm{diam}(\Omega) = \max_{\mathbf{x},\mathbf{y}\in\Omega} \|\mathbf{x} - \mathbf{y}\|_2.$$

Denote by $\mathbf{x}_s(t)$ the shock-interface location (a curve or surface in Ω). For a fixed radius fraction $r = 0.1$, the shock neighborhood is

$$\Omega_s \;=\; \big\{(\mathbf{x}, t) \in \Omega \times [0, T] \;\big|\; \|\mathbf{x} - \mathbf{x}_s(t)\|_2 \le r\, W\big\}.$$

Let $\mathcal{R}_t \subset \Omega$ be the rarefaction fan region at time t. Its space–time closure is

$$\Omega_r \;=\; \overline{\bigcup_{t\in[0,T]} \big(\mathcal{R}_t \times \{t\}\big)}.$$

H. Probconserv models

Together with Probconserv models (Hansen et al., 2024), we evaluate our method in the same setting on the 1D inviscid Burgers' equation where:

$$\partial_t u + u\,\partial_x u = 0,$$

defined on the spatial domain $x \in [-1, 1]$ and time interval $t \in [0, 1]$. The initial condition is

$$u(x, 0) = \begin{cases} -ax, & x < 0, \\ 0, & x \ge 0, \end{cases}$$

and Dirichlet boundary conditions are enforced: $u(-1, t) = a$ and $u(1, t) = 0$. The parameter $a > 0$ controls the initial shock strength. In our experiments, we test on both $a = 1$ (moderate rarefaction) and $a = 3$ (strong shock).

Figure A10 compares the mean squared error (MSE) over time for CI-PINN and probabilistic baselines on two regimes: param=1 (rarefaction-dominated) and param=3 (shock-dominated). CI-PINN consistently achieves the lowest global MSE in both settings, with approximately half the error of the best-performing competitor. Notably, in the simpler param=1 case, CI-PINN maintains a smooth and low MSE throughout the time horizon, demonstrating its strong generalization in smooth regimes.

In the shock-dominated case (param=3), CI-PINN exhibits small-amplitude, high-frequency oscillations in MSE after the shock forms. This pattern suggests that while the model may slightly misalign with the exact shock location at individual time steps, it quickly corrects itself—indicating stable tracking behavior. In contrast, other models accumulate error over time, especially near discontinuities. Overall, CI-PINN exhibits both low error and resilient behavior in the presence of shocks.

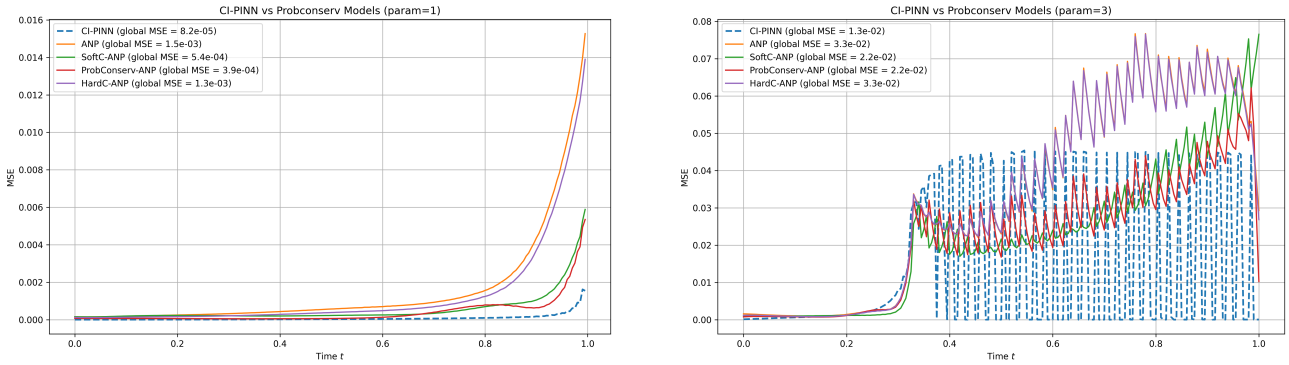


Figure A10. MSE by Time

I. Introduction and formulation of benchmark equations

I.1. Burgers' Equation

The Burgers' equation is a fundamental partial differential equation classified as a convection–diffusion equation. It plays a significant role in various fields of applied mathematics, including fluid mechanics, gas dynamics, and traffic flow(Souren Misra & Bobba, 2010; Hopf, 1950). The equation is expressed as:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \tag{25}$$

where $u = u(x,t)$ represents the velocity field, t denotes time, x is the spatial coordinate, and ν is the viscosity coefficient. When $\nu = 0$, the equation simplifies to the inviscid Burgers' equation, which can develop sudden changes (shock) in the solution. These shock arise because the characteristic wave speeds of this equation are given by $f'(u) = u$. To understand how shocks (sharp jumps) or rarefaction waves (smooth spreading) form, consider a point in the initial condition where u takes one value on the immediate left (call it u_L) and another value on the immediate right (call it u_R). If $u_L > u_R$, then the wave speed on the left side is higher than on the right side. The faster-moving characteristics from the left eventually overtake the slower ones on the right, causing the solution to "pile up" into a shock. Conversely, if $u_L < u_R$, then the wave speed on the left is lower than on the right. The characteristics spread out or "fan apart," resulting in a rarefaction wave (LeVeque, 1992a). A simple way to illustrate both cases is to use a square-wave initial condition. We initialize the solution with a piecewise-constant profile, setting $u(x,0) = 1$ for $x \in (-0.5, 0)$ and $u(x,0) = 0$ elsewhere.

Here, there are two transition points in the initial data (near $x = -0.5$ and $x = 0$). Depending on which side has the larger or smaller value of u, one observes either a shock or a rarefaction wave in the solution, see Figure 2.

It is also important to emphasize that the higher error values observed for methods such as Modified MLP, RWF, Causal, and Adaptive lr do not necessarily indicate poor performance. On the contrary, as illustrated in Appendix F, several of these methods exhibit promising behavior by closely approximating the shock structure.

I.2. Buckley-Leverett Equation

The displacement of two immiscible fluids is a significant challenge in the study of fluid flow within porous media. This phenomenon is often modeled by a PDE known as the Buckley-Leverett (BL) equation(Muskat, 1938). The BL equation provides a mathematical framework for studying the dynamics and interactions of these fluids under varying conditions in porous substrates. It is particularly relevant in secondary oil recovery, where water is injected into underground rock formations to displace additional unrecovered oil. The Buckley-Leverett equation is given by: $u_t + f(u)_x = 0$, where

$$f(u) = \frac{u^2}{u^2 + M \times (1-u)^2}.$$

Here, $u(x,t)$ represents the saturation of the injected fluid, and M is the mobility ratio, defined as the ratio of the mobility of the displacing fluid to the mobility of the displaced fluid. Since $f(u)$ is a non-convex function in the range $[0,1]$, the solution of the B-L equation exhibits both shocks and rarefaction waves, which are connected.

As an example, we consider $M = \frac{1}{4}$. The initial condition is prescribed as a piecewise-constant state with $u(x,0) = 1$ for $x \in [-1, -0.5]$ and $u(x,0) = 0$ for $x \in (-0.5, 1]$. visualized of clips are in the leftmost panel of Appendix F.2:

.

I.3. Euler System

The Euler equations of gas dynamics form a system of hyperbolic partial differential equations that govern the motion of an inviscid fluid. Derived from the fundamental principles of conservation of mass, momentum, and energy, these equations are extensively used in fluid dynamics and aerodynamics to model gas behavior (Toro, 2013). The Euler equations also serve as a benchmark for evaluating the accuracy of numerical methods.

The Euler system is expressed as: $\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \mathbf{F} = 0$, where

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ u(E+p) \end{pmatrix},$$

and $E = \frac{1}{2}\rho u^2 + \frac{p}{\gamma - 1}$.

Here, ρ denotes the density, u the velocity, p the pressure, E the total energy, and γ the ratio of specific heats. We benchmark our method on two canonical Euler shock-tube problems—Sod's and Lax's tests.The Sod and Lax shock tube problems are classical Riemann benchmarks that are widely used to assess the accuracy and shock-capturing robustness of numerical methods for the Euler equations (Sod, 1978; Lax, 1954; Toro, 2009).

The strong-form PINN baseline is omitted in the main text, as previous sections have shown its poor performance on conservation laws; full error tables, including that baseline, are reported in Appendix E.

I.3.1. Sod Problem

The Sod shock tube problem considers an ideal gas in a one-dimensional tube initially separated by a diaphragm, with the left state having higher density and pressure and the right state having lower density and pressure, while both sides are initially at rest (Sod, 1978). Upon removal of the diaphragm, the system evolves under the Euler equations and generates a sequence of characteristic waves; we initialize the system with a left state $(\rho, u, p) = (3,\ 0,\ 3)$ for $x \in [0, 0.5]$ and a right state $(\rho, u, p) = (1,\ 0,\ 1)$ for $x \in (0.5, 1]$.

I.3.2. Lax Problem

The Lax problem is a classical Riemann benchmark featuring a strong shock and a pronounced contact discontinuity; we initialize the system with a left state $(\rho, u, p) = (0.445,\ 0.698,\ 3.528)$ for $x \in [0, 0.5]$ and a right state $(\rho, u, p) = (0.5,\ 0,\ 0.571)$ for $x \in (0.5, 1]$. Owing to the coexistence of a strong shock and a contact discontinuity, the solution has limited regularity, rendering strong-form residual objectives ill-posed near discontinuities and making this problem a stringent test for PINN-based solvers

I.4. Two-Dimensional Euler Equations

The 2D Euler equations describe the dynamics of a compressible, inviscid fluid in two spatial dimensions. This system is governed by the conservation of mass, momentum, and energy (Li, 2002). The conservative form is given by:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = 0, \tag{26}$$

where the state vector $\mathbf{U}$ and the flux vectors $\mathbf{F}, \mathbf{G}$ are:

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho u v \\ u(E+p) \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \rho v \\ \rho u v \\ \rho v^2 + p \\ v(E+p) \end{pmatrix}. \tag{27}$$

Here, p is the pressure, ρ is the density, (u, v) are velocity components, and E is the total energy. The system is closed by the ideal gas law $p = (\gamma - 1)(E - \frac{1}{2}\rho(u^2 + v^2))$. 2D Euler problems are highly challenging for neural solvers due to the formation of oblique shocks and Mach stems.

We define initialize the system with:

$$(\rho, u, v, p)(x, y, 0) = \begin{cases} (2.0, 0.75, 0.5, 1.0) & \text{if } x < 0.5, y > 0.5 \text{ (LT)}, \\ (1.0, 0.75, -0.5, 1.0) & \text{if } x > 0.5, y > 0.5 \text{ (RT)}, \\ (1.0, -0.75, 0.5, 1.0) & \text{if } x < 0.5, y < 0.5 \text{ (LB)}, \\ (3.0, -0.75, -0.5, 1.0) & \text{if } x > 0.5, y < 0.5 \text{ (RB)}. \end{cases} \tag{28}$$

I.5. Shallow-Water Equations (SWE)

The Shallow-Water Equations are derived by depth-averaging the Navier-Stokes equations, assuming the horizontal length scale is much larger than the vertical depth (Vreugdenhil, 1994). They are widely used for modeling tsunami propagation and atmospheric flows. The 2D conservative form is:

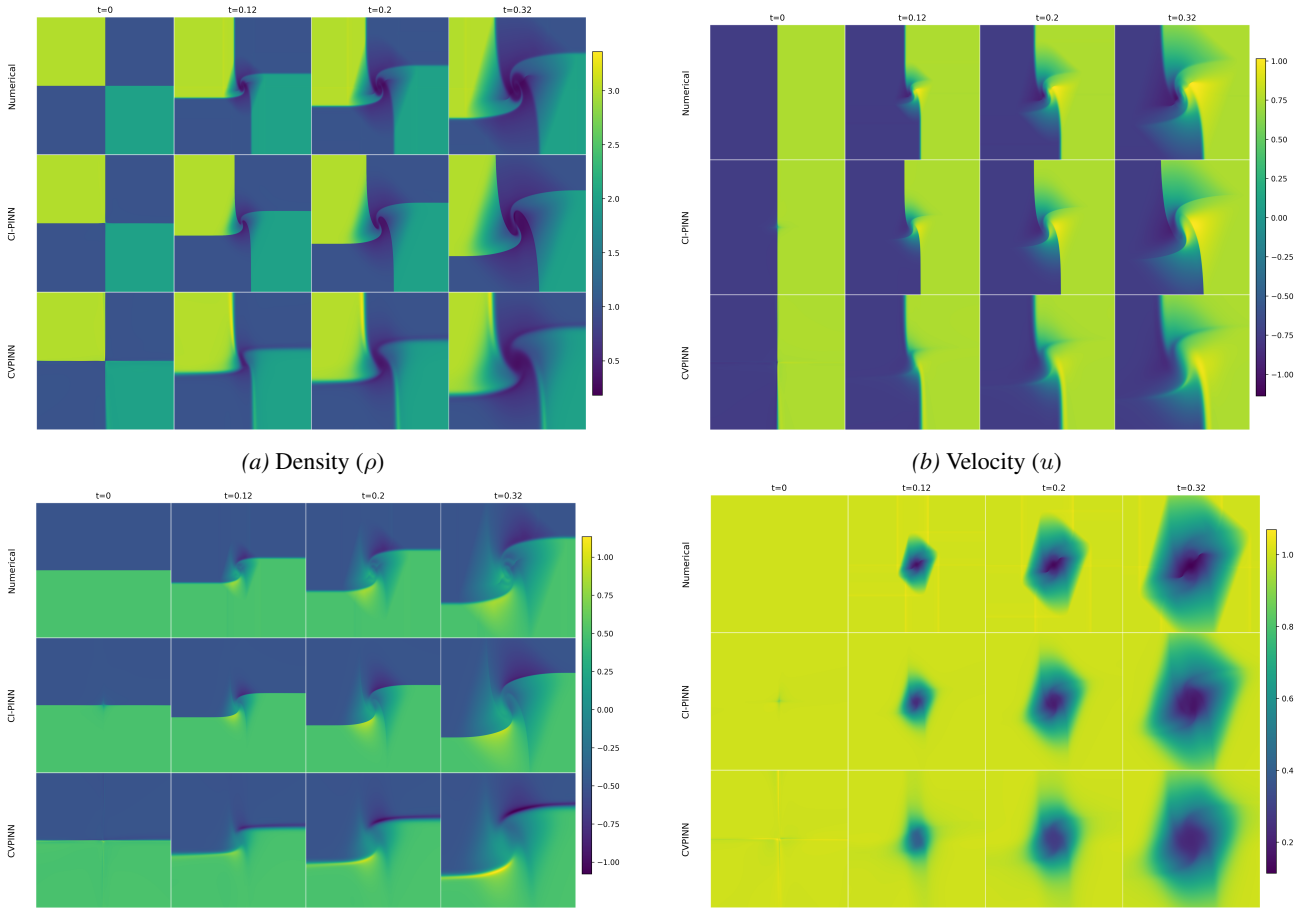
$$\frac{\partial}{\partial t}\begin{pmatrix} h \\ hu \\ hv \end{pmatrix} + \frac{\partial}{\partial x}\begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ huv \end{pmatrix} + \frac{\partial}{\partial y}\begin{pmatrix} hv \\ huv \\ hv^2 + \frac{1}{2}gh^2 \end{pmatrix} = 0, \tag{29}$$

where h is the fluid depth, g is the acceleration due to gravity, and (hu, hv) are the momentum components in the x and y directions, respectively.

I.5.1. Problem Setup

To evaluate shock-capturing performance in the presence of hydraulic jumps, we simulate a radial dam break scenario. The fluid depth is initialized with a discontinuous profile: $h = 2$ within a central circle of radius 0.5, and $h = 1$ in the surrounding domain.

J. 2D experiments



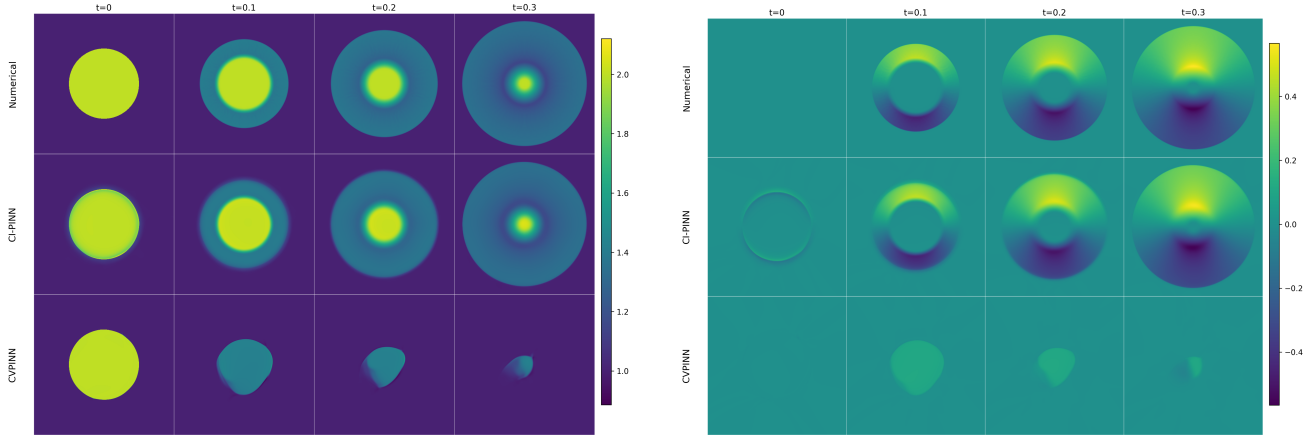
(a) Density (ρ)

(b) Velocity (u)

(c) Velocity (v)

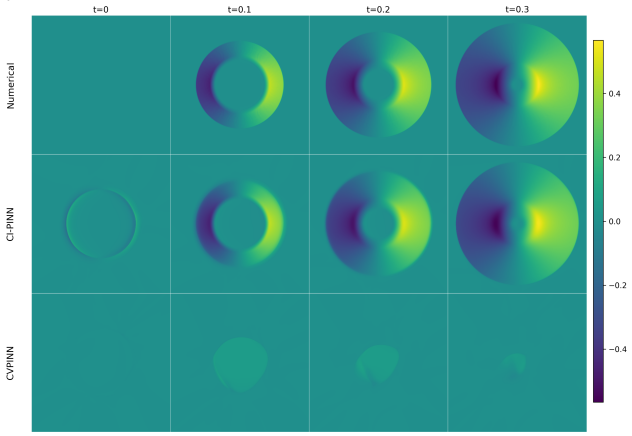
(d) Pressure (p)

Figure A11. 2D Euler system



(a) Density (h)

(b) Velocity (u)



(c) Velocity (v)

Figure A12. Shallow-Water Equations